Robust portfolio optimization

ROBUST
PORTFOLIO OPTIMIZATION

ROBUST PORTFOLIO OPTIMIZATION

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht
op gezag van de Rector Magnificus, Prof. Mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen, in het openbaar te verdedigen
op vrijdag 25 juni 2004 om 14.00 uur

door

Frank Johannes Willem Lutgens

Promotoren:
Prof. dr. ir. A. Kolen
Prof. dr. P.C. Schotman

Beoordelingscommissie:
Prof. dr. C.G. Koedijk (voorzitter)
Prof. dr. ir. D. den Hertog (Universiteit van Tilburg)
Dr. J.-P. Urbain

ISBN 90-9018143-1

Preface

Upon completion of this research, we almost reached the stadium which we originally intended as the point of departure for this research: multi-period portfolio choice problems. When I started my research in October 1999, my supervisors Antoon Kolen and Peter Schotman gave me the freedom to choose a subject for my dissertation, provided that it would be related to decision theory as well as finance. Apart from my main subject decision theory, the field of econometrics has always fascinated me. Not the least because I have always distrusted econometric models in their role to support decision making. This lead my interest to decision making under uncertainty, in particular robust optimization. Fortunately, finance is an area where an abundance of uncertainty is available. Already after two years I could meticulously describe the theme of my research as robust optimization applied to some problem in finance.

I am still pleased with this choice. The robust paradigm shows to be a valuable tool for decision making under uncertainty and has lead to surprising results.

This dissertation would not have been possible without the help and encouragement of several people. My supervisor Peter Schotman has always enthusiastically encouraged me. He embarked on this research from a different perspective which has lead to fruitful cooperation. I thank him for sharing his financial and econometric knowledge. I am grateful to my supervisor Antoon Kolen for his encouragement and useful comments on previous drafts of this thesis. I heartily thank Jos Sturm for our cooperation and the many things he taught me, professionally as well as personally.

Although my colleagues where quite busy themselves, they could always spare a moment (or two) which led to a pleasant working atmosphere. I am grateful for that. Special thanks go to my young (ex-) colleagues, including Bart, Bart and Kim. I enjoyed our many extracurricular activities.

Mijn vrienden ben ik dankbaar voor hun vriendschap en morele steun. In het bijzonder wil ik Julien bedanken voor onze inspirerende vergezochte reizen en diepzinnige discussies onderweg. Renske, jou wil ik bedanken voor je geduld en begrip tijdens de laatste maanden van dit onderzoek.

Veruit mijn grootste dank gaat uit naar mijn familie, met name mijn ouders. Hoewel jullie (net als ik) niet altijd precies wisten waar ik mee bezig was, toonden jullie altijd begrip wanneer ik in gedachten verzonken was. Zonder jullie steun en begrip had ik het niet gered.

Frank Lutgens
Maastricht, January 2004

Contents

Chapter 1

Introduction

Risk is the known chance of loss,
uncertainty is what is unknown.
Frank H. Knight (1885-1972)

Yesterday's portfolio theory was born in the 1950s with an innovative approach to investment developed by Professor Harry Markowitz. His approach surpasses traditional asset management, which focuses on predicting individual stock price movements using fundamental or technical analysis, by recognizing cross-sectional relations between assets and expressing the performance of a portfolio as a combination of its components' risk and return. It forms the theoretical foundation of a concept that practitioners and academics have long been aware off: "do not put all your eggs in one basket". Indeed, diversification reduces risk in the unsteady world. According to Markowitz, the optimal diversification ensues from a trade off between risk and return.

For the development of this operational theory for multidimensional portfolio choice under uncertainty, Markowitz was a joint Nobel Laureate for economics in 1990. His work has won wide acclaim due to its algebraic simplicity and, with the advent of computers in the 1970s to handle the vast number of calculations and ranges of historical data needed by the approach, suitability for empirical applications. Portfolio management today proceeds from the concepts brought up by Markowitz and melds theory and technology to optimize portfolio performance.

The crucial information needed for a practical implementation of modern portfolio theory concerns the behavior of future returns. Unfortunately, return behavior is not entirely understood and practitioners rely on approximate models to describe return behavior. They must choose a model specification and the numerical values for the parameters in that approximate model. Lack of consensus about the proper model specification and a wide collection of methods to provide numerical values for the parameters in the model give rise to a wide variety of return models. Alternative return models produce different optimal asset allocations. Accordingly, the quality of the asset allocation depends on the appropriateness of the return model and uncertainty about the aptness of a return model produces uncertainty about the optimal asset allocation.

Uncertainty about the return model emanates from model uncertainty and estimation uncertainty. *Model uncertainty* denotes potential model misspecification, that is, the selected return model may not be fully consistent with the true stochastic structure of the asset returns. The main source of model uncertainty is the need to make possibly false assumptions on the structure of the return model. Given a return model, *estimation uncertainty* refers to uncertainty about the proper values of some model parameters. The main source of estimation uncertainty is the limited amount of data available to infer parameter estimates. This causes the point estimates for the parameters, even in correctly specified models, to deviate from their true values.

Michaud (1989), Jobson and Korkie (1981, 1980) and Jorion (1986), among others, show that disregarding uncertainty leads to poor and unstable portfolio allocations. The Markowitz approach appears a reputed example that illustrates the effect of disregarding uncertainty. Michaud (1998) describes the Markowitz system as "a convenient and useful theoretical framework for portfolio optimality" but argues that "in practice it is an error prone procedure that often results in error maximized and investment-irrelevant portfolios". The problem is that small perturbations from a reference return model, ubiquitous in an uncertain environment, might lead to drastically different optimal asset allocations. In the Markowitz system this fault presents itself as the estimated return model generally reflects a deceptive positive bias on the expected portfolio performance. Therefore extreme positive or negative portfolio allocations are chosen that actually have a suboptimal performance which is much lower than the expected performance. A further complication is that perturbed return models are empirically indistinguishable from the reference model, and therefore it is unclear which asset allocation is best or at least reasonable.

Economists argue whether uncertainty should be treated differently from risk. According to Knight (1921), risk refers to situations where the decision maker can assign mathematical probabilities to the randomness which he is faced with. In contrast *Knightian uncertainty* refers to situations where the randomness cannot be quantified by mathematical probabilities. As John Maynard Keynes expressed it:

> "By 'uncertain' knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty...The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence...About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know." (J.M. Keynes, 1937)

The distinction between risk and uncertainty suggests that uncertainty should be treated differently from risk. This thesis studies the eclectic decision maker who is unsure about which (source of) information is most adequate. The decision maker is equipped with an *uncertainty set* of alternative, return models. She wishes to consider all return models, but does not (know how to) assign mathematical probabilities to individual models in this uncertainty set. As an alternative she aims for a decision that produces a robust performance: She maximizes the worst case performance that is possible for the models in the uncertainty set, in particular the least favorable model.

The robust decision maker is confident that the proper return model is contained in the uncertainty set. Consequently she is confident that if an asset allocation has good performance for the least favorable return model, it will surely perform well for the proper return model. The critical aspect of robust decision making is the uncertainty set of alternative return models which the investor considers. If this set is large, the worst case model may portray deplorable performance and the investor will behave as if she is very pessimistic. On the other hand, if the investor has a strong belief in some reference model and the set of alternatives is small, the optimal decision is almost equal to the optimal decision that is based exclusively on the the reference model but is not tolerant to deviations from this reference model.

This thesis contributes by studying the application of the robust paradigm to practically relevant portfolio choice problems. We consider the modelling aspects related to robust decision making and develop robust optimization models. Using these robust optimization models, we submit the robust paradigm to an empirical test of performance. The results show that a significant gain in robustness can be attained at the expense of, if any, only a small decrease in average performance.

1.1 Outline

The remainder of this chapter provides an overview of the literature on the robust approach to uncertainty and theoretical background on robust optimization.

In chapter 2, we consider single-period portfolio choice in the presence of estimation uncertainty. In the context of mean-variance portfolio choice, uncertainty in the mean expected returns has, if not accounted for, considerable negative impact on performance. We show how to set up the related robust optimization model and provide analytical solutions for robust portfolio choice. An empirical simulation on the Fama & French dataset sorted on size and book-to-market attributes shows the value of a robust solution. The remaining input for mean variance portfolio choice concerns the covariance matrix. We also demonstrate a robust approach to uncertainty in the covariance matrix and provide numerical solutions.

As an extension to mean-variance optimization we consider a benchmark tracking problem. We develop the estimation robust benchmark tracking problem, and use numerical optimization to test the robust approach in an empirical simulation on the Fama & French dataset.

In chapter 3, we study an investor who is uncertain about the correct model specification. The robust investor consults multiple experts who hold alternative prior beliefs about the correct return model, and the investor chooses a portfolio which is robust to these alternative return model priors. We derive, analytically, the model-robust portfolio choice for some stylized examples of model uncertainty and show that the robust portfolio, which is designed for best worst-case performance, actually has good expected performance too. For general return model priors we can compute the robust portfolio choice numerically. We perform an empirical study based on the Fama & French dataset to test the portfolio performance associated with a robust approach to multiple return model priors.

In chapter 4 we study multi-period portfolio choice based on a return model which features predictability, but suffers estimation uncertainty. We consider buy-and-hold and dynamic portfolio allocation for return models with different extents of estimation uncertainty and compare the robust portfolio choice to known results of non-robust portfolio choice. Moreover we characterize the parameter configuration which constitutes the worst case return model for the robust solution and show that this worst case parameter configuration varies among horizons and initial states.

Chapter 5 considers robust portfolio choice with options. Using conic duality theory, we develop the necessary techniques to formulate a robust optimization model which can handle options. We illustrate the applicability of this approach on a benchmark tracking problem and report the results of an empirical experiment.

Chapter 6 concludes with some recommendations and directions of future research.

1.2 Related literature

Uncertainty has attracted a lot of attention in various fields of literature, to mention statistical, econometric, financial and optimization literature.

Statistical decision theory attempts to improve estimation. The performance of an estimator is measured by the expected loss in objective that arises if we adopt an estimate instead of the true value. Ideally the objective is inspired on the portfolio choice problem, though studies generally use a quadratic loss function which features larger losses as the uncertainty or the bias of the estimator increase.

This has produced 'biased-shrink' estimators, used by James and Stein (1961), Jobson, Korkie and Ratti (1979), Jobson and Korkie (1981), Ledoit (1994) and Jorion (1986), that have uniformly smaller loss than conventional estimators such as the maximum likelihood estimator. Another class of estimators concerns minimax estimators that minimize the maximum performance. James and Stein (1961) and Judge and Bock (1978) show that minimax estimators exist which dominate the conventional maximum likelihood estimators in terms of expected loss.

Huber (1977, 1981) studies robust estimation in the context of data contamination. Huber (1981) shows that maximum likelihood estimators are not robust to slight misspecification of the underlying distribution. In the context of decision making under uncertainty Cavadini, Sbuelz and Trojani (2001) show that mean variance portfolio rules based on robustly estimated opportunity sets lead to stable performance for *local* (distributional) deviations from a relevant return model.

These approaches produce improved estimates which may be embedded in the optimization, but considerations regarding uncertainty are limited to the estimation stage. Estimators that perform well for some quadratic loss function need not be good estimators if measured in terms of the objective for portfolio choice. Moreover, the optimization may (and typically will) exploit the deviations of the estimate from its proper value for optimal decision making. Consequently in the context of optimization the previous approaches do not relieve the error-maximization problem presented in the introduction. Aït-Sahalia and Brandt (2001) pursue the first matter by considering direct estimation of portfolio weights.

Another stream of literature tries to overcome the effects of uncertainty by changing the optimization model. Jagannathan and Ma (2003) show why imposing supplementary 'wrong' constraints in the optimization model helps. They connect constraints in the optimization model to constrained estimation of the return model. Although promising results have been obtained in this context, the approach is merely a treatment of symptoms by limiting the harm that uncertainty may do. Regrettably it also limits the positive potential of the approach by limiting the solution space by imposing 'wrong' constraints. Ter Horst, de Roon and Werker (2002) propose an adjustment to the coefficient of risk aversion that incorporates the estimation uncertainty for the mean returns. The adjustment follows from an analysis of the expected loss in utility due to maximum likelihood estimation. This approach produces results similar to a Bayesian approach.

Growing attention is directed at explicitly considering uncertainty in decision making. Black and Litterman (1990, 1992) combine the market information with additional views of the investor to reduce uncertainty. Similarly Bayesian approaches, studied by Kandel and Stambaugh (1996), Pàstor and Stambaugh (2000), Barberis (2000) and Pàstor (2000), manage uncertainty by a presumption of prior information on the return model. The prior information is combined with the observed data to form the predictive distribution of returns which features uncertainty. The predictive distribution serves as subjective distribution for decision making according to expected utility theory which was developed by Savage (1954).

Ellsberg (1961) describes a mind experiment that for the first time raised a serious objection to the expected utility paradigm of Savage (1954). The empirically observed preference ordering deduced from this mind experiment was inconsistent with *any* conceivable (even if subjective) probability measure through expected utility maximization.

Gilboa and Schmeidler (1989) suggest model uncertainty as an explanation to Ellsberg's

(1961) paradox: the decision maker has too little information to form a unique prior and considers a set of priors as conceivable. On account of the *behavioral assumption* of uncertainty aversion, the decision maker considers the minimal expected utility over all priors in the set when evaluating a decision. Various studies elaborated on the (multiple priors) *maximin* expected utility theory of Gilboa and Schmeidler (1989), for example Dow and Werlang (1992), Epstein and Wang (1994), Chen and Epstein (2002). Chamberlain (2000) applies the maximin expected utility theory to a situation where finitely many subjective priors are given for a parametric model.

Maximin expected utility theory relates to the recent literature on robustness, e.g. Anderson, Hansen and Sargent (1999), Maenhout (1999) and Uppal and Wang (2003), which has its roots in robust control theory. In this approach, the preference for robustness is translated to a stochastic game of a robust investor who anticipates the malevolent actions of her opponent, nature, which will select the worst case return model associated with the investor's decision. The statistical theory of detection is used to quantify the set of alternative model specifications that the investor should fear. The robust investor postulates a decision which has maximal worst case performance over this set. The size of the set reflects the investor's preference for robustness. Alternatively, the investor may express her preference for robustness by penalizing decisions according to their uncertainty in the objective. Hansen, Sargent, Turmuhambetova and Williams (2002) show that these approaches corresponding to using an uncertainty penalty or uncertainty set lead to similar results.

Cavadini et al. (2001) and Aït-Sahalia and Brandt (2001, p.1313) study robust portfolio choice for uncertainty which results from small (distributional) contaminations to some reference model. This type of uncertainty is common in the literature on ambiguity, e.g. Epstein and Wang (1994).

Although the formalizations of the preceding approaches, i.e. corresponding to Gilboa and Schmeidler (1989), Anderson et al. (1999) and Cavadini et al. (2001), differ, they build on the same motivation: the decision maker is uncertain about the true return model and aims for a solution which is robust to this uncertainty. One could argue that the approaches of Anderson et al. (1999) and Cavadini et al. (2001) correspond to specific parsimonious quantifications of the 'multiple priors' uncertainty set in Gilboa and Schmeidler (1989). Kogan and Wang (2002) and Uppal and Wang (2003) elaborate on these equivalent approaches. For a critical discussion on this literature, we refer to Sims (2001) and Pathak (2002).

In the literature on optimization, Ben-Tal and Nemirovski (1997, 1998) and El Ghaoui, Oustry and Lebret (1998) consider robust optimization which primarily concentrates on quantitative techniques to deal with uncertainty in accordance with the robust paradigm. More specifically, robust optimization studies (non-finite) uncertainty sets of candidate coefficients that can be handled efficiently. As for finance, Lobo (2000), Goldfarb and Iyengar (2003) and Costa and Paiva (2002) apply the approach to mean–variance portfolio choice under parameter uncertainty and Ben-Tal, Margalit and Nemirovski (2000) solve a multistage portfolio choice problem. One could say that robust optimization provides *maximin* expected utility theory with quantitative methods to manage practical relevant portfolio choice problems.

In this thesis we consider a decision maker who uses the robust paradigm common in the work of Gilboa and Schmeidler (1989), Anderson et al. (1999), Ben-Tal and Nemirovski (1997) and El Ghaoui et al. (1998). We frame the analysis of robust decision making in the jargon of robust optimization. We will base the uncertainty set of alternative optimization coefficients on economic and statistical arguments. In chapters 2 and 4, we derive uncertainty sets to account for estimation uncertainty. Model uncertainty will be added in chapter 3 by specifying multiple return model priors.

The methods we present are directed at solving practical relevant problems. We add to the literature on maximin expected utility theory by developing operational methods to handle relevant non-finite uncertainty sets. We can deal with uncertainty resulting from not necessarily small deviations from some reference model.

1.3 Robust optimization

Consider an optimization problem :

$$\max_x \{f(x, \zeta) : \; x \in X(\zeta) \subset R^N\} \tag{1}$$

where

- $x \in R^N$ is the decision vector, e.g. a vector of portfolio weights,
- $\zeta \in R^M$ denotes the problem instance by a vector of optimization coefficients, e.g. a vector of expected future returns, and
- the dimensions M, N and the mappings f, X are structural elements of the problem.

We study a decision environment for which the particular problem instance is uncertain: knowledge of ζ is confined to its membership of some uncertainty set $\mathcal{U}$ (for convenience we assume $\mathcal{U}$ is compact).

We consider the robust paradigm to handle uncertainty in the optimization problem. According to the robust paradigm, a solution is robust feasible if

$$x \in X(\zeta) \qquad \forall \zeta \in \mathcal{U}. \tag{2}$$

We denote the set of robust feasible solutions by

$$X_R = \bigcap_{\zeta \in \mathcal{U}} X(\zeta).$$

Moreover, a solution is robust optimal if, additionally, it provides the best possible worst case value

$$\min_{\zeta \in \mathcal{U}} f(x, \zeta)$$

of the original objective, i.e. it solves the deterministic *robust optimization problem*:

$$\max_x \{ \min_{\zeta \in \mathcal{U}} f(x, \zeta) : \; x \in \bigcap_{\zeta \in \mathcal{U}} X(\zeta) \} \tag{3}$$

For convenience we use an equivalent form for general robust optimization problems[1]

$$\max_x \{ f(x) : \; x \in \bigcap_{\zeta \in \mathcal{U}} X(\zeta) \} \tag{4}$$

On operational grounds we pursue to model our financial decision problem as an optimization problem which can be solved efficiently. A semiformal description of an efficiently solvable optimization problem is a problem which requires "not too many"[2] standard arithmetic operations to provide an optimal solution with any given precision. Optimization techniques that efficiently solve optimization problems are only available for special classes of optimization problems.

The fundamental division of efficiently solvable problems lies in convex versus non-convex optimization problems. A convex optimization problem is an optimization problem where we seek a minimum (maximum) of a convex (concave) function over a convex set. With a proof based on the ellipsoid method (Khachiyan (1979), Ben-Tal and Nemirovski (2001) show that under some 'mild' assumptions, convex optimization problems are efficiently solvable. One of these assumptions concerns polynomial computability: For a given solution we can evaluate the objective, find a sub-gradient of the objective, check feasibility and report a separating hyperplane in the case of infeasibility in "not too many" operations.

A robust optimization problem is convex if f is convex and the set of robustly feasible solutions X_R is convex (which holds at least if each $X(\zeta)$ is convex). Unfortunately, finite representable uncertainty sets with an infinite number of elements violate the assumption of polynomial computability as checking feasibility requires to verify $x \in X(\zeta)$ for infinitely many ζ. On the other hand if X is described by finitely many non-negativity restrictions

$$X(\zeta) = \{x : \; g_i(x, \zeta) \geq 0, \; i = 1, .., N\},$$

[1]Without loss of generality, the objective in problem (4) does not depend on ζ. Problems for which the objective depends on the problem instance ζ (i.e. $f(x, \zeta)$), may be cast in the form (4) by introducing an auxiliary variable x_0 and transformations $X(\zeta) \to \{(x_0, x) : x \in X(\zeta) \cap x_0 \leq f(x, \zeta)\}$ and $f(x) \to x_0$. As we maximize f and impose $x_0 \leq f(x, \zeta) \; \forall \zeta \in \mathcal{U}$, $x_0 = \min_{\zeta \in \mathcal{U}} f(x, \zeta)$.

[2]In number bounded by a polynomial of the problem size and a given precision of the solution.

we can check robust feasibility by considering

$$\min_{i=1,\dots,N} \min_{\zeta \in \mathcal{U}} g_i(x,\zeta) \geq 0. \tag{5}$$

A seperating hyperplane is also available as the sub-differential of g_{i^*}, i^* minimizes (5). Consequently if $\mathcal{U}$ is a convex set and problem (5) satisfies the 'mild' assumptions for efficiently solvability, we can solve a convex robust optimization problem efficiently.

Results of this tenor merely have theoretical relevance. The complexity bounds that can be established by the ellipsoid-method, are polynomial but "too large" for practical application.

Another, relatively new polynomial time method to solve "well-structured" generic convex optimization problems is the interior point method which was originally introduced by Karmarkar (1984) for linear optimization and extended to convex optimization by Nesterov and Nemirovski (1994) . For some special conic optimization problems , among which linear, conic quadratic and semi-definite optimization problems, the interior point method has, next to the best known polynomial complexity bound, extremely efficient practical performance. Moreover efficient interior-point software to solve conic problems is available (Sturm (1999), Anderson (2000), Alizadeh, Haeberly, Nayakkankuppam and Overton (1997)).

This narrows the target group of optimization models that we preferably use to model our financial decision problem to conic optimization problems.

The only aspect of the robust optimization problem that may still obstruct fast computations is its nested structure: each check of robust feasibility requires solving the optimization problem (5). Ideally we can transform this robust optimization problem to a non-nested conic optimization problem. This would be possible if we can transform the infinite number of restrictions (2) to an equivalent finite set of conic restrictions. Consider the problem

$$\begin{array}{ll} \max_{x_0,x} & x_0 \\ \text{subject to} & g(x,\zeta) \geq x_0 \qquad \forall \zeta \in \mathcal{U}. \end{array} \tag{6}$$

If $\mathcal{U}$ has an infinite number of elements, then (6) contains an infinite number of constraints. We are interested in a *transformation* which replaces the infinite number of constraints by a finite number. Possibly one can achieve this by introducing auxiliary variables $y \in R^p$ and finding constraints $h_j(x_0,x,y) \geq 0,\ j = 1,..,q$ such that the following relation holds for all x_0 and x:

$$g(x,\zeta) \geq x_0 \qquad \forall \zeta \in \mathcal{U} \ \Leftrightarrow\ \exists y \in R^p : h_j(x_0,x,y) \geq 0,\ j = 1,..,q. \tag{7}$$

In section 1.3.1 we demonstrate such a transformation and provide a short overview of other useful transformations published in recent literature.

Second order cone programming

Many problems in this thesis are (transformed to) second order cone optimization problems. An $(N+1)$-dimensional second order (Lorentz) cone SOC is defined as

$$SOC = \{(z_0, z) \in R_+ \times R^N : \sqrt{z_1^2 + z_2^2 + .. + z_N^2} \leq z_0\} = \{(z_0, z) : \|z\| \leq z_0\}.$$

The linear cone ($z_0 \geq 0$) is a special type of second order cone with $z = 0$. The second order cone belongs to the cone of positive semi-definite matrices as the positive semi-definite constraint

$$\begin{pmatrix} z_0 I_N & z \\ z' & z_0 \end{pmatrix} \succeq 0$$

holds, by the Schur complement, if and only if $(z_0, z) \in SOC$. These solid, pointed and closed convex cones are self-dual. This implies that the dual cone to C is

$$C^* = \{y \in R^{N+1} : z'y \geq 0 \; \forall z \in C\} \tag{8}$$

falls within the same class as the primal problem, for example $SOC^* = SOC$. A second order cone optimization problem (SOCP) is defined as

$$\begin{aligned} &\min \; c'x \\ &\text{subject to } A_i x - b_i \in SOC \; \forall i = 1, .., M \end{aligned} \tag{9}$$

The dual problem (SOCD) is defined as

$$\begin{aligned} &\max \sum_{i=1}^{M} b_i' y \\ &\text{subject to } \sum_{i=1}^{M} A_i' y_i = c \\ &\qquad y_i \in SOC \; \forall i = 1, .., M. \end{aligned} \tag{10}$$

Weak duality, complementary slackness and hence strong duality results hold for second order cone optimization problems so that problems (9) and (10) have, given a Slater condition, the same optimal value. Using interior point methods we can solve SOCPs (and also optimization problems with semi-definite constraints) in polynomial time.

Apart from linear constraints some quadratic and bilinear constraints can be modelled as (rotated) second order constraints:

$$\frac{1}{2}x'x \leq x_0 y_0, \; x_0 + y_0 \geq 0 \Leftrightarrow \left\| \begin{matrix} \frac{x_0 - y_0}{2} \\ x \end{matrix} \right\| \leq \frac{x_0 + y_0}{2} \Leftrightarrow \begin{pmatrix} \frac{x_0+y_0}{2} \\ \frac{x_0-y_0}{2} \\ x \end{pmatrix} \in SOC \tag{11}$$

This result is useful for expressing the variance of a multi-asset portfolio w with covariance matrix Σ:

$$var(w) = w'\Sigma w.$$

Let C be the lower triangular part of the Choleski decomposition of the covariance matrix, i.e. $\Sigma = CC'$ and let $x = A'w$. A constraint of the variance $w'\Sigma w \leq x_0$ suits a second order representation (11) with $y_0 = 1/2$.

1.3.1 Problem transformations

In this section we illustrate several techniques for transforming a robust feasible set which is described by an infinite number of constraints.The non-finite number of constraints results from imposing some constraint for all parameter configurations in some uncertainty set. We transform the feasible set to an equivalent description which involves a finite number of constraints. Such transformations are valuable for solving robust optimization problems but unfortunately only available for specific combinations of constraints and uncertainty sets. After the illustration we give a short overview of the constraint – uncertainty combinations for which a useful transformation is known.

Consider a feasible set described by a linear constraint

$$X(\zeta) = \{x : \ \zeta' x \geq x_0\}$$

with uncertainty parameters which are known to belong to an ellipsoidal uncertainty set[3]

$$\mathcal{U} = \{\zeta : \ (\zeta - \hat{\zeta})\Omega^{-1}(\zeta - \hat{\zeta}) \leq \theta^2\} = \{\hat{\zeta} + Cu \mid \|u\| \leq \theta\}.$$

Note that the uncertainty set contains an infinite number of elements. The robust feasible set,

$$X_R = \bigcap_{\zeta \in \mathcal{U}} X(\zeta) = \{x : \ \zeta' x \geq x_0 \ \forall \zeta \in \mathcal{U}\},$$

is thus described by an infinite number of linear constraints.

Ben-Tal and Nemirovski (1998) provide a transformation for affine constraints with uncertain coefficients that belong to an ellipsoid:

$$\begin{array}{rcll} \zeta' x & \geq & x_0 & \forall \zeta \in \mathcal{U} = \{\hat{\zeta} + Cu \mid \|u\| \leq \theta\} \\ (\hat{\zeta} + Cu)' x & \geq & x_0 & \forall u : \ \|u\| \leq \theta \\ \hat{\zeta}' x + u' C' x & \geq & x_0 & \forall u : \ \|u\| \leq \theta \end{array}$$

[3] $\Omega = CC'$ is positive definite, $(\zeta - \hat{\zeta})'\Omega^{-1}(\zeta - \hat{\zeta}) = (\zeta - \hat{\zeta})C^{-T}C^{-1}(\zeta - \hat{\zeta}) = u'u$, hence the equivalent representation of the ellipsoid with $u = C^{-1}(\zeta - \hat{\zeta})$.

Using the *Cauchy-Schwarz inequality* and $\|u\| \leq \theta$,

$$\begin{aligned} \hat{\zeta}'x + u'C'x &\geq \hat{\zeta}'x - \|u\|\|C'x\| \\ &\geq \hat{\zeta}'x - \theta\|C'x\| = \hat{\zeta}'x - \theta\sqrt{x'CC'x}. \end{aligned} \tag{12}$$

Inequality (12) is tight for $u = -\theta\frac{C'x}{\|C'x\|}$. Consequently,

$$\zeta'x \geq x_0 \ \forall\zeta \in \mathcal{U} \qquad \Leftrightarrow \qquad \hat{\zeta}'x - \theta\|C'x\| \geq 0 \tag{13}$$

This is a second order cone (SOC) constraint and may be included in efficiently solvable second order cone optimization problems (SOCP).

We illustrate two alternative techniques to derive useful problem transformations: optimality conditions and duality.

Consider the problem

$$\min_{\zeta}\{\zeta'x : \ \zeta \in \mathcal{U}\}. \tag{14}$$

The Lagrangian is

$$L(\zeta) = \zeta'x - \lambda(\zeta - \hat{\zeta})'(CC')^{-1}(\zeta - \hat{\zeta}),$$

and the *optimality conditions* are

$$\begin{aligned} x - \lambda(CC')^{-1}(\zeta - \hat{\zeta}) &= 0, \qquad \lambda \geq 0 \\ (\zeta - \hat{\zeta})'(CC')^{-1}(\zeta - \hat{\zeta}) - \theta^2 &= 0 \end{aligned}$$

Substituting the solution to the optimality conditions, $\zeta = \hat{\zeta} + \theta\frac{CC'x}{\sqrt{x CC'x}}$, into the constraint $\zeta'x \geq x_0$ produces (13). We apply this technique in chapter 2 to deduce a robust feasible set for benchmark optimization problems.

Yet another approach is to use *conic duality theory* to characterize the cone of linear functions $\zeta'x - x_0$ that are nonnegative on the set $\mathcal{U}$, i.e.

$$\{(-x_0, x) : \ \zeta'x - x_0 \geq 0, \ \forall\zeta \in \mathcal{U}\} = \{(-x_0, x) : \ \begin{pmatrix} 1 \\ \zeta \end{pmatrix}' \begin{pmatrix} -x_0 \\ x \end{pmatrix} \geq 0, \ \forall \begin{pmatrix} 1 \\ \zeta \end{pmatrix} \in \tilde{\mathcal{U}}\}. \tag{15}$$

with $\tilde{\mathcal{U}} = \{(1, \zeta) : \ \zeta \in \mathcal{U}\}$. Note the similarity between (15) and conic duality (8). The only difference is that $\tilde{\mathcal{U}}$ is a conic section instead of a cone. Indeed

$$\tilde{\mathcal{U}} = \{(1, \zeta) : \ \begin{pmatrix} \theta \\ C^{-1}(\zeta - \hat{\zeta}) \end{pmatrix} \in SOC\} = \{(1, \zeta) : \ \begin{pmatrix} \theta & 0' \\ -C^{-1}\hat{\zeta} & C^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ \zeta \end{pmatrix} \in SOC\}.$$

As the second order cone is self dual (see (8)),

$$\begin{pmatrix} t \\ v \end{pmatrix}' \begin{pmatrix} \theta & \\ -C^{-1}\hat{\zeta} & C^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ \zeta \end{pmatrix} \geq 0$$

whenever $(t, v) \in SOC$. Hence the robust feasible set (15) is equivalent to

$$\{x : \begin{pmatrix} -x_0 \\ x \end{pmatrix} = \begin{pmatrix} \theta & \\ -C^{-1}\hat{\zeta} & C^{-1} \end{pmatrix}' \begin{pmatrix} t \\ v \end{pmatrix}, \begin{pmatrix} t \\ v \end{pmatrix} \in SOC\}$$

which is equivalent to (13). This technique is applied in chapter 5.

A related approach is to use *programming duality*. By strong duality, the optimal value of (14) is larger than x_0 if and only if the optimal value of the dual is larger than x_0. If we model (14) as the primal (9), the dual of the form (10) implies

$$\begin{aligned} C^{-T}\lambda &= x && (16) \\ -\theta\lambda_0 + \hat{\zeta}'C^{-T}\lambda &\geq x_0 && (17) \\ (\lambda_0, \lambda) &\in SOC && (18) \end{aligned}$$

If we use (16) to substitute for λ and subsequently use (18) in (17), the feasible set (13) follows.

Other useful transformations have recently been reported. Ben-Tal and Nemirovski (1998) show that linear (affine) constraints combined with an uncertainty set characterized by a finite number of second order cone restrictions can be transformed to a finite number of conic quadratic constraints. A derivation for robust versions of affine constraints over uncertainty sets which result from the intersection of an ellipsoid and a polyhedral set is given in chapter 5 of this thesis.

Quadratic constraints combined with polyhedral uncertainty sets are generally NP hard to solve (Luo, Sturm and Zhang 2001). Yakubovich's (1977) $\mathcal{S}$-lemma implies that a homogenous quadratic constraint with an uncertainty set defined by a homogenous quadratic constraint, reduces to linear matrix inequalities which may be part of an efficiently solvable semi-definite optimization problem. Ben-Tal and Nemirovski (1998) and El Ghaoui et al. (1998) generalize the $\mathcal{S}$-lemma to non-homogeneous, conic (convex) quadratic functions. Ben-Tal and Nemirovski (2001) and Sturm and Zhang (2003) derive linear matrix inequalities to quantify the robust feasible set of quadratic constraints over a domain defined by a quadratic inequality. Sturm and Zhang (2003) also report results for uncertainty sets given by an equality constraint in a strictly convex or concave quadratic function, or the combination of a convex quadratic and linear inequality. *Quadratic constraints* under uncertainty sets characterized by the intersection of ellipsoids are generally NP hard.

Uncertain semi-definite constraints with general ellipsoidal uncertainty lead to robust counterparts that are NP hard to solve (Luo et al. 2001). Ben-Tal and Nemirovski (1998) propose some "well-structured" ellipsoidal uncertainty sets and provide a transformation, for which the results are tractable. El Ghaoui et al. (1998) consider semi-definite constraints for which the data matrices are special rational functions of the uncertain coefficients and, for a special uncertainty set, provide sufficient (semi-definite) conditions for robust optimality. Luo et al. (2001) consider special quadratic matrix inequalities over uncertainty sets, among which the ellipsoidal uncertainty set, defined by special quadratic matrix inequalities and show that the robust constraint is a linear matrix inequality.

Ben-Tal and Nemirovski (1998) and Ben-Tal, Nemirovski and Roos (2001) approximate the uncertainty set by some form for which an efficient transformation is known. The robustness of the solution depends on the approximate uncertainty set: (i) expanding the uncertainty set results in conservative solutions, (ii) shrinking the uncertainty set results in less robust solutions, and (iii) distorting the uncertainty set may affect the solution either way.

Approximations to the robust optimization problem that result from finite approximations to the uncertainty set are typically inadequate. Accurate approximations require many elements and induce large optimization problems. Less accurate approximations that produce a manageable problem size typically loose robustness and are only adequate for very "smooth" optimization problems.

Notes on robust optimization problems

Robust optimization problems are connected to maximin optimization problems. The robust problem (3) in standard form is a special maximin problem with a feasibility set that depends on the chosen ζ. However an asymmetric treatment of the constraints, $x : x \notin X(\zeta)$ are infeasible but $\zeta : x \notin X(\zeta)$ imply unbounded solutions, prevents minimax results[4].

It is important to be careful when modelling robustness in *multi-stage* decision problems. One-stage decision problems require that all decisions are taken before uncertainty is resolved. A robust decision is a *here-and-now* decision which satisfies each separate constraint for every coefficient in the uncertainty set. Alternatively multi-stage problems allow part of the decisions to be taken after (some) uncertainty is resolved. Consequently constraints may not be treated separately. We illustrate this with an example:

$$\min_{x,z_1,z_2} \{-x + z_1 + z_2 : x \in [1,2],\ -z_1 \leq \zeta x \leq z_2\ \forall \zeta \in \{-1,1\}\ \}. \tag{19}$$

If (19) is a one stage problem, robust feasible solutions satisfy $z_1,\ z_2 \geq |x|$ and the

[4]Omit the constraint by defining

$$g(x,y) = \begin{cases} f(x) & \text{if } x \in X(\zeta) \\ -\infty & \text{otherwise} \end{cases}$$

and consider the equivalent maximin problem (with a feasible set independent of ζ),

$$\max_x \min_{\zeta\in\mathcal{U}} g(x,\zeta).$$

A minimax result is not available as g violates the necessary condition of convexity ζ for each x.

robust optimum is 1 at $(x, z_1, z_2) = (1, 1, 1)$. Alternatively, let (19) be a two-stage problem, and let $z_1(\zeta)$ and $z_2(\zeta)$ be second stage decisions that are to be taken after uncertainty is resolved and may exploit the new information ζ. In this case, relations between constraints play a role for the first stage decision x: If $\zeta = -1$, the constraints require $-z_2(-1) \leq x \leq z_1(-1)$, for $\zeta = 1$ these require $-z_1(1) \leq x \leq z_2(1)$. For either case, one inequality is not tight. The corresponding slack can be used by second stage decisions to improve the objective. The optimal decision $x = 2$ has robust objective value 2.

We can clarify the distinction between one-stage and multi-stage explicit by stressing the timing of decisions in (19), i.e.

$$\min_{x\in[1,2]} \max_{\zeta\in\{-1,1\}} \left\{ \min_{z_1,z_2} -x + z_1 + z_2 : \; -z_1 \leq \zeta x \leq z_2 \right\}. \tag{20}$$

1.4 Shaping the uncertainty set

Crucial input for the robust optimization is the specification of the uncertainty set. On the one hand it must provide an adequate description of the uncertainty; on the other hand it should permit a mathematical characterization that allows for efficiently computable solutions (see section 1.3.1).

A description of the uncertainty set is adequate if it is the smallest set of return models that is trusted to contain the proper return model. Any addition to the uncertainty set may result in conservatism as a robust decision accounts for the additional, though redundant restraint.

The decision maker will determine the basis for the uncertainty set.

Ben-Tal and Nemirovski (1998) argue that ellipsoids and intersections of ellipsoids produce valuable type of uncertainty sets. Ellipsoids can represent relations between uncertain data elements, which is an improvement over a typically conservative representation of uncertainty by a Carthesian product of uncertainty sets of individual uncertain optimization coefficients. Ben-Tal and Nemirovski (1998) appeal to a universal though coarse argument, often applied in engineering, that a random variable is almost never worse than three standard deviations from its mean. This argument finds theoretical backing by (non-parametric) estimates such as the Chebychev inequality. Other stochastic arguments are the ellipsoidal distributions which have ellipsoidal confidence regions. The intersections of ellipsoids may be used to approximate a wide variety of more complicated confidence regions.

A Bayesian foundation to shape the uncertainty set is also compatible with Knightian uncertainty: Bayesians combine a *subjective* prior with the characteristics of the data (expressed by a likelihood function) to form a posterior distribution. In a natural way the highest posterior density region can serve as uncertainty set of (subjectively) plausible models and the cumulative density, e.g. set to 95%, of this region serves as a measure for robustness.

A frequential approach applied to estimation uncertainty, uses the finite sample distribution as basis for construction of the uncertainty set. The finite sample distribution replaces the posterior distribution in the Bayesian approach, and proceeds similar henceforth. A frequentist approach does not rely on subjective priors, yet is not compatible with Knightian uncertainty.

In this thesis we generally follow the Bayesian approach to form the uncertainty set, although the treatment of estimation uncertainty is also supported by frequentist arguments. We deviate slightly from the described Bayesian approach as regards the the highest posterior density property of the uncertainty set. Computational restraints (section 1.3.1) urge us to consider deviating regions, nonetheless with equivalent extent of robustness measured by cumulative probability.

Chapter 2

Estimation uncertainty[1]

As far as the laws of mathematics refer to reality,
they are not certain,
and as far as they are certain,
they do not refer to reality.
Albert Einstein (1879-1955)

In this chapter we study the implications of a robust approach to parameter uncertainty in mean-variance portfolio choice. The work is motivated by the difficulty in estimating accurately the stochastic process of the returns. Jobson and Korkie (1981), Jorion (1986) and Michaud (1998), among others, show that disregarding the associated uncertainty leads to poor and unstable portfolio allocations. The Markowitz (1952) model is a reputed example that illustrates the effect of disregarding the uncertainty. Michaud (1998) describes the Markowitz system as "a convenient and useful theoretical framework for portfolio optimality" but argues that "in practice it is an error prone procedure that often results in error maximized and investment-irrelevant portfolios". The problem is that small perturbations to a reference return model, ubiquitous in an uncertain environment, might lead to drastically different mean-variance portfolios.

Robust portfolio choice aims at designing portfolios with a performance that is robust to alternative plausible return models. Although the robust paradigm concedes to lose against a strategy tuned to *maximize* expected performance, it wins in terms of *reliability* of actually achieving the robust expected performance. Moreover it is conceivable that the robust portfolio actually leads to superior ex-post performance. When small errors in the estimates lead to drastically non-optimal mean-variance portfolios, a robust portfolio might, in terms of ex-post performance, be preferred to a naive portfolio which maximizes performance without considering uncertainty.

In this chapter we confine our analysis to single period mean-variance portfolio choice on multiple risky assets. For special forms of uncertainty we derive analytical solutions

[1]This chapter is based on joint work by Frank Lutgens and Peter Schotman.

to robust portfolio choice. We demonstrate that a robust investor's portfolio choice is sensitive to alternative formulations of the mean-variance problem which produce the same portfolio choice without uncertainty. We will also consider portfolio choice when risk is measured relative to a benchmark.

We conclude with an empirical test of the robust approach which shows that a robust approach leads to reliable portfolio choice, solves the error-maximization critique and may have ex-post expected performance that is superior to the naive mean-variance portfolio performance.

2.1 Portfolio choice under certainty

We consider the problem of allocating wealth over N different risky assets and a riskfree asset. The riskfree asset has return r_f. Expected returns of the risky assets, in excess of the riskfree rate, are denoted by the N-vector μ. The $(N \times N)$ covariance matrix of returns is denoted by Σ. A portfolio is an N-vector w with elements w_i which are the fractions of wealth allocated to the risky asset i. The balance of wealth, $1 - \iota' w$, is invested in the riskfree asset. In general, if we impose restrictions on portfolio weights, we will describe a set $\mathcal{W} \subset R^N$ and require $w \in \mathcal{W}$.

We begin with a review of the basic mean-variance theory. In the standard Markowitz model the investor faces a tradeoff between the expected return and variance of a portfolio. The investor's objective function, interpreted as expected utility, is

$$(1 - \iota' w) r_f + w'(\mu + \iota r_f) - \frac{1}{2}\gamma w' \Sigma w \tag{1}$$

where γ is a measure of risk aversion. Since the riskfree rate does not affect the investment decision in this problem, we subtract r_f from the objective function. The mean variance problem (MV) is to maximize

$$Q(w) = \mu' w - \frac{1}{2}\gamma w' \Sigma w. \tag{2}$$

In this section we assume that μ and Σ are known. In this case, the first order optimality condition for the concave objective function is

$$\mu - \gamma \Sigma w = 0. \tag{3}$$

and hence the optimal solution is

$$w_\gamma = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{4}$$

with optimal value $Q(w_\gamma) = \frac{1}{2\gamma}\mu' \Sigma^{-1} \mu$.

An alternative way to characterize optimal mean-variance portfolios is to select the portfolio that has maximum expected utility for a given variance σ^2. This leads to the feasible set $\mathcal{W}_\sigma = \{w : w'\Sigma w \leq \sigma^2\}$ and the variance constrained mean problem (M),

$$\max_{w \in \mathcal{W}_\sigma} \mu' w. \tag{5}$$

The first order optimality conditions are

$$\begin{aligned} \mu - \lambda \Sigma w &= 0, \qquad \lambda \geq 0 \\ w' \Sigma w &= \sigma^2 \end{aligned} \tag{6}$$

and the optimal solution is

$$w_\sigma = \frac{\sigma}{\sqrt{\mu' \Sigma^{-1} \mu}} \Sigma^{-1} \mu. \tag{7}$$

with optimal value $\sigma \sqrt{\mu' \Sigma^{-1} \mu}$. The optimal solution is equal to the solution of the MV problem if

$$\sigma = \frac{\sqrt{\mu' \Sigma^{-1} \mu}}{\gamma}. \tag{8}$$

For every level of risk aversion γ there is a corresponding volatility constraint σ. Without parameter uncertainty the two approaches are equivalent. However, since the relation between σ and γ depends on μ, estimation uncertainty in μ will destroy the equivalence between (2) and (5).

The dual to maximizing expected return is minimization of the variance for a given level of expected return. Let $\bar{\mu} > 0$ and define $\mathcal{W}_\mu = \{w : w'\mu \geq \bar{\mu}\}$. The mean constrained variance problem (V) is

$$\min_{w \in \mathcal{W}_\mu} w' \Sigma w. \tag{9}$$

The first order optimality conditions are

$$\begin{aligned} \Sigma w - \lambda \mu &= 0, \qquad \lambda \geq 0 \\ \mu' w &= \bar{\mu} \end{aligned} \tag{10}$$

and the optimal solution is

$$w_\mu = \frac{\bar{\mu}}{\sqrt{\mu' \Sigma^{-1} \mu}} \Sigma^{-1} \mu, \tag{11}$$

with optimal value $\frac{\bar{\mu}}{\mu' \Sigma^{-1} \mu}$. Without parameter uncertainty, the optimal solution is equivalent to the solution of the MV problem if $\bar{\mu} = \frac{\sqrt{\mu' \Sigma^{-1} \mu}}{\gamma}$. But also in this case the equivalence breaks down when there is parameter uncertainty as the relation between $\bar{\mu}$ and γ then depends on uncertain μ. Moreover, when μ is uncertain and we use an estimate $\hat{\mu}$ instead, the optimal portfolio based on $\hat{\mu}$ might actually violate the true constraint $\mu \in \mathcal{W}_\mu$ when $\hat{\mu}$ deviates from μ.

A fourth manifestation of the mean–variance problem is the maximization of the Sharpe ratio (SR),

$$\frac{\mu' w}{\sqrt{w'\Sigma w}}. \tag{12}$$

We use the convention that the Sharpe ratio associated with $w = 0$ is zero. We assume $\mu \neq 0$ and consider the equivalent optimization problem

$$\begin{aligned} &\max_{Sh,w} Sh \\ &\mu' w - Sh\sqrt{w'\Sigma w} \geq 0 \end{aligned} \tag{13}$$

with first order optimality conditions

$$\begin{aligned} 1 - \lambda\sqrt{w'\Sigma w} = 0, &\qquad \lambda \geq 0 \\ \lambda(\mu' w - Sh\sqrt{w'\Sigma w}) = 0 & \end{aligned} \tag{14}$$

which are satisfied (and define a maximum) for

$$w_{Sh} = \alpha\Sigma^{-1}\mu \tag{15}$$

for every positive scalar α. All previously considered optimal portfolios also attain the maximum Sharpe ratio $\rho = \sqrt{\mu'\Sigma^{-1}\mu}$. The optimal Sharpe portfolio (15) determines the relative weights of the risky assets in the portfolio, but not the leverage, i.e. the total amount invested in the risky assets. As estimation uncertainty affects the expected returns and risk of the risky assets, but not the riskfree rate r_f, the leverage of the portfolio determines the exposure to estimation uncertainty. The Sharpe ratio is insensitive to leverage and is therefore not the most appropriate measure to evaluate portfolio performance under uncertainty.

The final formulation of the mean-variance problem compares expected utility Q of the optimal decision with the expected utility of suboptimal portfolios. Expected utility of the optimal decision w_γ follows from substituting (4) in (2):

$$Q(w_\gamma) = \frac{\rho^2}{2\gamma}. \tag{16}$$

The loss in utility due to selecting some portfolio w different from the optimal portfolio is defined by the loss function

$$L(w) \;=\; Q(w_\gamma) - Q(w). \tag{17}$$

Hence

$$\begin{aligned} 2\gamma L(w) &= \mu'\Sigma^{-1}\mu - 2\gamma\mu' w + \gamma^2 w'\Sigma w \\ &= (\mu - \gamma\Sigma w)'\Sigma^{-1}(\mu - \gamma\Sigma w). \end{aligned} \tag{18}$$

Obviously the optimal portfolio w_γ has zero loss. Without parameter uncertainty minimizing expected loss is equivalent to maximizing mean–variance utility (2). With parameter uncertainty the target $Q(w_\gamma)$ is itself uncertain. Behaviorally, an investor who minimizes loss is concerned with the relative performance compared to an uncertain benchmark.

Benchmark tracking constrained portfolio choice

Many institutional investors define risk relative to some benchmark portfolio. An example is the tracking error which is the difference between a benchmark portfolio $\tilde{w}$ of risky assets and a portfolio w chosen by the investor. The benchmark portfolio has an implicit investment of $(1 - \iota'\tilde{w})$ in the riskfree asset and a return $r_f + \tilde{w}'y$. The investor's portfolio has return $r_f + w'y$ as before. A typical constraint on the tracking error is

$$\mathsf{E}\left[(w - \tilde{w})'(yy')(w - \tilde{w})\right] \leq \tau^2$$

or equivalently by using the moments,

$$(w - \tilde{w})'(\Sigma + \mu\mu')(w - \tilde{w}) \leq \tau^2. \tag{19}$$

We consider the problem (BT) of maximizing expected return subject to a tracking error constraint,

$$\max_{w \in \mathcal{W}_B} \quad \mu' w \tag{20}$$

with $\mathcal{W}_B$ denoting the set of portfolios w that satisfy (19).

The first order optimality conditions require

$$\begin{aligned} \mu - \lambda(\Sigma + \mu\mu')(w - \tilde{w}) &= 0 \\ w &\in \mathcal{W}_B. \end{aligned} \tag{21}$$

Solving for w and using the inversion lemma $(\Sigma + \mu\mu')^{-1} = \Sigma^{-1} - \frac{\Sigma^{-1}\mu\mu'\Sigma^{-1}}{1+\mu'\Sigma^{-1}\mu}$ give the optimal solution

$$w_\tau = \tilde{w} + \frac{\tau(\Sigma + \mu\mu')^{-1}\mu}{\sqrt{\mu'(\Sigma + \mu\mu')^{-1}\mu}} = \tilde{w} + \frac{\tau}{\rho\sqrt{1+\rho^2}}\Sigma^{-1}\mu. \tag{22}$$

The tradeoff between the tracking error and expected return is parameterized by the maximum tracking error τ. When we increase the maximum allowed tracking error τ, the optimal investment is drawn towards the return maximizing portfolio. The optimal portfolio is a combination of the benchmark portfolio and the optimal mean-variance portfolio.

2.2 Robustness to unknown mean

In practice μ and Σ are not known, but are estimated from a sample of T observations y_t of historical returns. The sensitivity of the portfolio on small perturbations in the coefficients, combined with the uncertainty in the estimator for μ, make the *naive* solution, which simply ignores uncertainty, "non-robust". This section reconsiders mean-variance analysis for the case that expected returns are estimated, but the variance is known.

Let $\hat{\mu}$ be an estimator of μ and let $\Omega = \mathsf{E}[(\hat{\mu}-\mu)(\hat{\mu}-\mu)']$ describe the uncertainty around μ. For example, $\hat{\mu}$ and Ω could be the posterior mean and covariance matrix of μ. We consider a set of plausible expected returns,

$$\mathcal{U} = \{\mu : (\mu-\hat{\mu})'\Omega^{-1}(\mu-\hat{\mu}) \le \theta^2\} \tag{23}$$

with θ the desired degree of robustness. The uncertainty set conveys the intuition that deviations from $\hat{\mu}$ in directions $\mu - \hat{\mu}$ for which uncertainty is large, measured by Ω, are more plausible than deviations from μ in directions for which uncertainty is small. In typical applications θ is set such that $\mathcal{U}$ contains p of the probability mass around $\hat{\mu}$. For example if returns are normally distributed, the set $\mathcal{U}$ may be parameterized by a Chi-square distribution with N degrees of freedom and θ is set equal to the pth percentile of the Chi-square distribution with N degrees of freedom, i.e. $\theta^2 = \chi^2_{inv}(p, N)$. We assume that both $\hat{\mu}$ and Ω are given to the investor.

The naive approach to mean–variance portfolio choice is to ignore uncertainty and substitute $\hat{\mu}$ for μ to obtain the portfolio $\hat{w} = \alpha\Sigma^{-1}\hat{\mu}$ with α depending on the particular version of the mean–variance problem being solved. This portfolio promises more than can be expected. For instance, the expectation of the squared optimal Sharpe ratio is

$$\begin{aligned}
\mathsf{E}[\hat{\rho}^2] = \mathsf{E}[\hat{\mu}'\Sigma^{-1}\hat{\mu}] &= \mathsf{E}[\mathsf{tr}(\Sigma^{-1}\hat{\mu}\hat{\mu}')] \\
&= \mathsf{E}[\mathsf{tr}(\Sigma^{-1}(\mu\mu' + \Omega))] \\
&= \rho^2 + \mathsf{tr}(\Sigma^{-1}\Omega)
\end{aligned} \tag{24}$$

Since Σ and Ω are both positive definite matrices, the second term in (24) is positive. The promised optimal Sharpe ratio $\hat{\rho}^2$ is therefore an upward biased estimate of what an investor can actually expect: the true expected value ρ^2. Given (24) we can determine the expected utility,

$$\begin{aligned}
\mathsf{E}[Q(\hat{w}_\gamma)] &= \mathsf{E}[\hat{w}_\gamma'\mu - \tfrac{1}{2}\gamma\hat{w}_\gamma'\Sigma\hat{w}_\gamma] \\
&= \mathsf{E}[\tfrac{1}{\gamma}\hat{\mu}'\Sigma\mu - \tfrac{1}{2\gamma}\hat{\mu}'\Sigma\hat{\mu}] \\
&= \tfrac{1}{2\gamma}(\rho^2 - \mathsf{tr}(\Sigma^{-1}\Omega))
\end{aligned} \tag{25}$$

and the expected value of the loss (17) is

$$\begin{aligned}
\mathsf{E}[L(\hat{w}_\gamma)] &= \mathsf{E}[Q(w_\gamma) - Q(\hat{w}_\gamma)] \\
&= \tfrac{1}{2\gamma}\left(\rho^2 - (\rho^2 - \mathsf{tr}(\Sigma^{-1}\Omega))\right) \\
&= \tfrac{1}{2\gamma}\mathsf{tr}(\Sigma^{-1}\Omega)
\end{aligned} \tag{26}$$

The bias will typically depend on the number of assets, N, and the precision of the estimator Ω. An important special case is $\hat{\mu} = \bar{y}$, the sample mean of the observed data y_t, and $\Omega = \Sigma/T$. This arises if the prior on μ is uninformative. In this case expected utility (25) will only be positive if $\rho^2 > N/T$ and the expected loss is $\frac{1}{2\gamma}N/T$. Hence, in the presence of uncertainty, a naive approach may suffer possibly large expected loss and is therefore not necessarily the optimal portfolio choice.

We continue with a study of the robust approach to mean-variance portfolio choice. If the weighting of mean against variance is given by the risk aversion parameter γ, the robust expected utility optimization problem (RMV) is

$$\max_{w} \min_{\mu \in \mathcal{U}} \mu' w - \frac{1}{2}\gamma w' \Sigma w. \tag{27}$$

From equation (13) in chapter 1 we know that the robust expected mean is

$$\min_{\mu \in \mathcal{U}} \mu' w = \hat{\mu}' w - \theta \sqrt{w' \Omega w}. \tag{28}$$

Hence the robust problem (RMV) reduces to

$$\max_{w} \hat{\mu}' w - \theta \sqrt{w' \Omega w} - \frac{1}{2}\gamma w' \Sigma w. \tag{29}$$

A robust investor will invest if and only if the robust expected return is positive. Such a portfolio exists if and only if

$$\max_{w} w' \hat{\mu} - \theta \sqrt{w' \Omega w} \geq 0. \tag{30}$$

The left-hand side is equivalent to

$$\begin{aligned}
&\max_{w} \; w'\hat{\mu} - \theta\sqrt{w'\Omega w} \\
&\max_{s \geq 0} \; \max_{w: w'\Omega w = s^2} \; w'\hat{\mu} - \theta s \\
&\max_{s \geq 0} \left(\max_{w: w'\Omega w = s^2} w'\hat{\mu} \right) - \theta s \\
&\max_{s \geq 0} \sqrt{\hat{\mu}\Omega^{-1}\hat{\mu}} s - \theta s
\end{aligned} \tag{31}$$

where the last implication follows from (7). The optimal value is positive if and only if

$$\hat{\mu}\Omega^{-1}\hat{\mu} > \theta^2. \tag{32}$$

In statistical terms the robust investor will only consider risky assets if $\hat{\mu}$ is significantly different from zero using a χ^2-test of size $1 - p$. When the preference for robustness θ and the uncertainty matrix Ω are large ($\sqrt{\hat{\mu}\Omega^{-1}\hat{\mu}} \leq \theta$), the robust investor omits risky investment since the uncertainty penalty $\theta\sqrt{w'\Omega w}$ will outweigh the expected return.

When $\Omega = \Sigma/T$, (32) presents a test on the estimated Sharpe ratio $\hat{\rho} = \sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}$. If the Sharpe ratio is large compared to uncertainty $1/T$ and the investor's preference for robustness, more precisely when $\hat{\rho} \geq \theta/\sqrt{T}$, the investor is confident that the true Sharpe ratio is different from zero and therefore finds it attractive to invest[2].

Empirically the preliminary test (32) is not redundant. Campbell, Lo and Mackinley (1997, p. 206) suggest a baseline typical sample Sharpe ratio of $\hat{\mu}/\sigma = 8\%/20\% = 0.4$. In this case, a robust mean-variance investor with 5 years of historical data, who perceives any returns within $\theta = 2$ standard errors from the mean as plausible will stay out of the market. An excess expected return of 8% estimated with only five years of data is not enough to generate a positive robust excess return. Only if the robust investor has more data or a strong prior belief in a positive equity premium, she has enough confidence to participate in the market.

The first order conditions for the robust problem (RMV) are

$$\hat{\mu} - \left(\frac{\theta}{\sqrt{w'\Omega w}}\Omega - \gamma\Sigma\right) w = 0 \tag{33}$$

This is a nonlinear system that we are unable to solve analytically for general Ω. The numerical solution is straightforward though as the problem is an second order cone optimization problem (SOCP). Note that (29) is equivalent to

$$\max_w \{z : z \leq w'\hat{\mu} - \theta\sqrt{w'\Omega w} - \frac{1}{2}\gamma w'\Sigma w\}. \tag{34}$$

Let A and B be the lower triangular parts of the Choleski decompositions of Σ and Ω. By equation (11) in chapter 1, (34) is equivalent to

$$\max_w \{z : \begin{pmatrix} w'\hat{\mu} - z - \frac{1}{2}\gamma\sigma \\ B'w \end{pmatrix} \in SOC, \begin{pmatrix} \frac{1}{2}(\sigma^2+1) \\ \frac{1}{2}(\sigma^2-1) \\ A'w \end{pmatrix} \in SOC\}.$$

This is a SOCP and therefore it can be solved efficiently by available software implementations of interior point methods, for example Sturm (1999). For the special case the solution is summarized in the next theorem.

Theorem 1 *Assume $\Omega = \Sigma/T$. The optimal portfolio of the robust, expected utility,*

[2]One could argue that, if the true Sharpe ratio is small, the robust investor enters the market only when the sample error is sufficient to produce a large Sharpe ratio and will therefore suffer considerable losses. To resolve this paradox we need to consider on what information the investor conditions: The investor does not know the true Sharpe ratio but conditions on the sample. From this perspective, upon observing a high sample Sharpe ratio, more likely than not, the true Sharpe ratio is also high.

mean variance problem (29) *is*

$$\begin{aligned} w_r &= \left(1 - \frac{\theta}{\hat{\rho}\sqrt{T}}\right)\frac{1}{\gamma}\Sigma^{-1}\hat{\mu} && \textit{if} \quad \hat{\rho}\sqrt{T} \geq \theta \\ &= 0 && \textit{otherwise}, \end{aligned} \tag{35}$$

Proof From (32) it follows that the robust portfolio is zero if $\hat{\rho}\sqrt{T} < \theta$. Let us assume $\hat{\rho}\sqrt{T} \geq \theta$. Then the robust problem (RMV) is equivalent to

$$\begin{aligned} &\max_{w} w'\hat{\mu} - \frac{\theta}{\sqrt{T}}\sqrt{w'\Sigma w} - \frac{1}{2}\gamma w'\Sigma w = \\ &\max_{s\geq 0} \max_{w:w'\Sigma w=\sigma^2} w'\hat{\mu} - \frac{\theta}{\sqrt{T}}\sigma - \frac{1}{2}\gamma\sigma^2 = \\ &\max_{s\geq 0} \sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}\sigma - \frac{\theta}{\sqrt{T}}\sigma - \frac{1}{2}\gamma\sigma^2 = \\ &\frac{1}{2\gamma}\left(\sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}} - \frac{\theta}{\sqrt{T}}\right)^2 \end{aligned} \tag{36}$$

for optimal $\sigma = \frac{1}{\gamma}\left(\sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}} - \frac{\theta}{\sqrt{T}}\right)$ and corresponding optimal portfolio

$$w = \frac{\sigma}{\sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}}\Sigma^{-1}\hat{\mu} = \left(1 - \frac{\theta}{\hat{\rho}\sqrt{T}}\right)\frac{1}{\gamma}\Sigma^{-1}\hat{\mu}$$

□

The robust expected utility investor only enters the market if the estimated returns are significantly different from zero. Moreover the robust utility investor reduces the exposure to the risky portfolio by a factor $\frac{\theta}{\hat{\rho}\sqrt{T}}$, which is between zero and one and is therefore always strictly more conservative than a naive mean variance investor.

Theorem 1 also shows the distinction between the treatment of risk and uncertainty. A robust investor deals with risk aversion γ by investing more or less in the bundle of risky assets that has maximal Sharpe ratio. However, irrespective of the risk aversion, the robust investor may choose to be remain passive if uncertainty is large.

The robust version of the variance constrained mean problem (RM) is

$$\max_{w\in\mathcal{W}_\sigma} \min_{\mu\in\mathcal{U}} w'\mu \tag{37}$$

Using (28) to solve the inner minimization, the robust problem (RM) reduces to

$$\max_{w\in\mathcal{W}_\sigma} w'\hat{\mu} - \theta\sqrt{w'\Omega w}. \tag{38}$$

For the same reasons as for the expected utility problem (RMV), we are unable to solve the general robust problem (RM) analytically. However also this problem presents a second order cone optimization problem which can be solved efficiently. For the special case $\Omega = \Sigma/T$ an analytic solution is available.

Theorem 2 *Assume $\Omega = \Sigma/T$. The optimal portfolio of the robust, variance constrained, mean variance problem (37) is*

$$\begin{aligned} w_{\sigma,r} &= \tfrac{\sigma}{\hat{\rho}}\Sigma^{-1}\hat{\mu} && \text{if } \theta \leq \hat{\rho}\sqrt{T} \\ &= 0 && \text{otherwise,} \end{aligned} \tag{39}$$

with $\hat{\rho} = \sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}$.

Proof When a portfolio w has positive robust expected mean, then this also holds for the portfolio αw with $\alpha = \sigma/\sqrt{w'\Sigma w}$ which satisfies the variance constraint. Hence from (32) it follows that the robust portfolio is zero and $w \in \mathcal{W}_\sigma$ is not binding if $\hat{\rho}\sqrt{T} < \theta$. Let us assume $\hat{\rho}\sqrt{T} \geq \theta$. Then the robust problem (RM) is equivalent to

$$\max_{w: w'\Sigma w = \sigma^2} \; w'\hat{\mu} - \frac{\theta}{\sqrt{T}}\sigma = \sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}\sigma - \frac{\theta}{\sqrt{T}}\sigma$$

for $w = \frac{\sigma}{\sqrt{\hat{\mu}\Sigma^{-1}\hat{\mu}}}\Sigma^{-1}\hat{\mu}$ □

Figure 2.1 summarizes the results for the special case. The naive portfolios are equal for a suitable choice of γ and σ (see (8)) but the robust portfolios differ. A robust expected utility (RMV) investor always invests less than her naive colleague. A robust mean (RM) investor, if active, holds the same portfolio as her naive colleague but recognizes that the robust expected portfolio return is lower than the naive expected portfolio return.

In the special case with $\Omega = \Sigma/T$, no new information is added as uncertainty, like risk, is characterized by covariance matrix and therefore produces the same directions for optimal portfolio choice as a naive mean-variance analysis. Alternatively if uncertainty diverges from risk, i.e. Ω is not proportional to Σ, directions with large uncertainty are not necessarily equal to directions of large risk and therefore a robust portfolio does not necessarily have small risk and vice versa. In that case, the robust portfolio follows from an optimal tradeoff between the portfolio return, the variance constraint and the preference for robustness.

The robust version of the mean constrained variance problem (RV) is

$$\min_{w}\{w'\Sigma w : \mu' w \geq \bar{\mu} \; \forall \mu \in \mathcal{U}\} \tag{40}$$

Note that the constraint is equivalent to

$$\min_{\mu \in \mathcal{U}} \mu' w \geq \bar{\mu} \Leftrightarrow \hat{\mu}' w - \theta\sqrt{w'\Omega w} \geq \bar{\mu} \tag{41}$$

where the last implication follows from (32).

For the same reasons as for the robust problem (RMV), we are unable to solve the general robust problem (RV) analytically. However also this problem presents a second order cone optimization problem which can be solved efficiently. As for the previous robust portfolios, an analytic solution to (RV) is available for the special case $\Omega = \Sigma/T$:

$$\begin{aligned} w_{\mu,r} &= \left(\hat{\rho} - \tfrac{\theta}{\sqrt{T}}\right) \tfrac{\hat{\mu}}{\hat{\rho}} \Sigma^{-1} \hat{\mu} && \text{if} \quad \theta \leq \hat{\rho}\sqrt{T} \\ &= 0 && \text{otherwise,} \end{aligned} \tag{42}$$

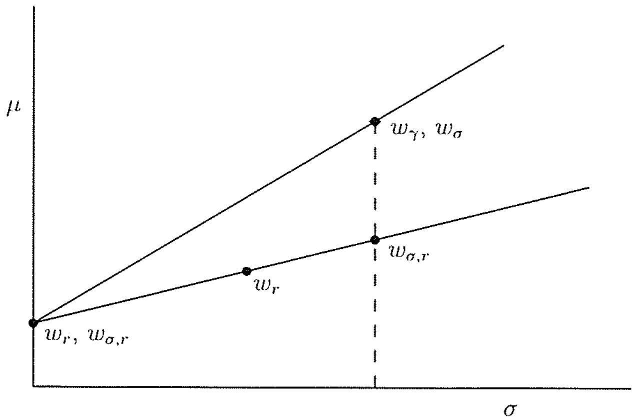
The robust version of the Sharpe ratio problem (RSR) is

$$\max_{w} \min_{\mu \in \mathcal{U}} \frac{\mu' w}{\sqrt{w' \Sigma w}}, \tag{43}$$

hence by (12) and (32) equivalent to

$$\begin{aligned} &\max_{Sh,w} Sh \\ &\quad \mu' w - \theta \sqrt{w' \Omega w} - Sh \sqrt{w' \Sigma w} \geq 0 \end{aligned} \tag{44}$$

Figure 2.1: Optimal portfolios of robust mean variance investors



Notes: The figure shows the optimal portfolio decisions for three types of investors in (standard deviation, expected portfolio return) space. All investors divide their wealth between the market portfolio and the riskfree asset. The naive mean-variance investors (w_γ, w_σ) are on the estimated Capital Market Line. The robust capital market line has a smaller slope due to estimation uncertainty in expected return. A robust mean investor ($w_{\sigma,r}$) with a variance constraint either chooses the same portfolio (for which she expects a lower return though) or invests solely in the riskfree asset. A robust expected utility investor (w_r) invests less in the market portfolio, and will invest solely in the riskfree asset under the same conditions as the robust mean investor.

Note that the optimal robust solution to the robust problem (RSR) remains scalable. For the special case $\Omega = \Sigma/T$, any solution to the previous robust problems (RMV), (RM) and (RV) is optimal.

Benchmark tracking constrained portfolio choice

For benchmark tracking constrained problems, the uncertainty about the expected returns affects the objective function as well as the constraint set $\mathcal{W}_B$.

The robust portfolio optimization problem requires that $w \in \mathcal{W}_B$ holds for all $\mu \in \mathcal{U}$. In particular it must hold for the worst expected return $\mu \in \mathcal{U}$ which will depend on the portfolio w evaluated. For a particular w the worst μ for the tracking constraint is given by the solution of

$$\max_{\mu \in \mathcal{U}} (w - \tilde{w})'(\Sigma + \mu\mu')(w - \tilde{w}). \tag{45}$$

Solving for μ, we obtain the robust tracking error constraint.

Theorem 3 *The set $\bigcap_{\mu \in \mathcal{U}} \mathcal{W}_B$ of robust feasible portfolios*

$$\mathcal{W}_{Br} = \{w : \ (w - \tilde{w})'\Sigma(w - \tilde{w}) + (\|\hat{\mu}'(w - \tilde{w})\| + \|\theta A'(w - \tilde{w})\|)^2 \leq \tau^2\} \tag{46}$$

with A' as the upper triangular part of the Choleski decomposition of Ω such that $\Omega = AA'$.

Proof Use

$$\{\mu : \ (\mu - \hat{\mu})\Omega^{-1}(\mu - \hat{\mu}) \leq \theta^2\} = \{\mu : \ \mu = \hat{\mu} + Au, \ \|u\| \leq \ \theta\}$$

to substitute for μ in (45):

$$\max_{\|u\| \leq \theta} (w - \tilde{w})'(\Sigma + \hat{\mu}\hat{\mu}' + 2\hat{\mu}u'A' + Auu'A')(w - \tilde{w}). \tag{47}$$

Define $p = A'(w - \tilde{w})$ and consider the relevant terms in (47),

$$2\hat{\mu}'(w - \tilde{w})u'p + p'uu'p \tag{48}$$

Use $u'p = \cos\gamma \ \|u\|\|p\|$ with $\gamma \in [0, \pi]$ as the angle between u and p to substitute for $u'p$ in (48),

$$2\hat{\mu}'(w - \tilde{w}) \cos\gamma\|u\|\|p\| + \cos^2\gamma\|u\|^2\|p\|^2. \tag{49}$$

For our estimation we distinguish between two cases.

(i) $\hat{\mu}'(w - \tilde{w}) \geq 0$. In this case the following holds:

$$\begin{aligned} 2\hat{\mu}'(w-\tilde{w})\cos\gamma\|u\|\|p\| & +\cos^2\gamma\|u\|^2\|p\|^2 \\ & \leq 2\hat{\mu}'(w-\tilde{w})\|u\|\|p\| + \|u\|^2\|p\|^2 \\ & \leq 2\hat{\mu}'(w-\tilde{w})\theta\|p\| + \theta^2\|p\|^2 \end{aligned} \tag{50}$$

Hence (50) provides an upperbound on (49) which is attained by $u^* = \theta\frac{p}{\|p\|}$, hence tight.

(ii) $\hat{\mu}'(w - \tilde{w}) < 0$. In this case the following holds:

$$\begin{aligned} 2\hat{\mu}'(w-\tilde{w})\cos\gamma\|u\|\|p\| & +\cos^2\gamma\|u\|^2\|p\|^2 \\ & \leq -2\hat{\mu}'(w-\tilde{w})\|u\|\|p\| + \|u\|^2\|p\|^2 \\ & \leq -2\hat{\mu}'(w-\tilde{w})\theta\|p\| + \theta^2\|p\|^2 \end{aligned} \tag{51}$$

Hence (51) provides an tight upperbound on (49) which is attained by $u^{**} = -\theta\frac{p}{\|p\|}$.

Theorem (3) follows by substituting u^* and u^{**} in (47) and taking the maximum. □

The robust benchmark portfolio choice problem is:

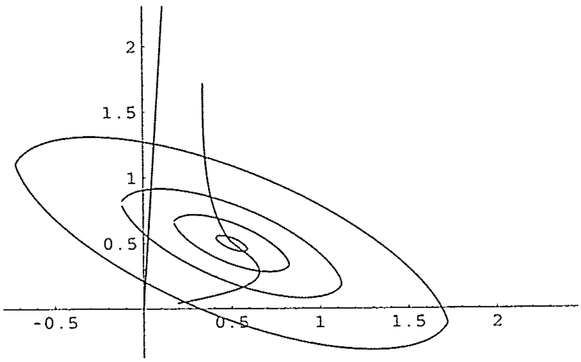
$$\max_{w\in\mathcal{W}_{Br}} \ \hat{\mu}'w - \theta\sqrt{w'\Omega w} \tag{52}$$

This is again a second order cone optimization problem. Let A and B be the lower triangular parts of the Choleski decompositions of Ω and Σ. Consider z_1 and z_2 such that $\|\theta A'(w-\tilde{w})\| \leq z_1$, $\|\hat{\mu}'(w-\tilde{w})\| \leq z_2$ and $(w-\tilde{w})'\Sigma(w-\tilde{w}) + (z_1+z_2)^2 \leq \tau^2$. For any feasible solution $w \in \mathcal{W}_{Br}$, we can find z_1 and z_2 such that the inequalities hold ($z_1 = \|\theta A'(w-\tilde{w})\|$ and $z_2 = \|\hat{\mu}'(w-\tilde{w})\|$). Conversely any solution (z_1, z_2, w) to the inequalities presents a feasible solution $w \in \mathcal{W}_{Br}$. Moreover, the equalities are second order cone constraints. Therefore problem (52) is equivalent to the SOCP

$$\begin{aligned} & \max z_0 \\ & \begin{pmatrix} \hat{\mu}'w - z_0 \\ \theta A'w \end{pmatrix} \in SOC \\ & \begin{pmatrix} \tau \\ z_1 + z_2 \\ B(w-\tilde{w}) \end{pmatrix} \in SOC \\ & \begin{pmatrix} z_1 \\ \theta A'(w-\tilde{w}) \end{pmatrix} \in SOC \\ & \begin{pmatrix} z_2 \\ \hat{\mu}'(w-\tilde{w}) \end{pmatrix} \in SOC. \end{aligned} \tag{53}$$

We are not able to derive analytical results for the general case, even for the special case $\Omega = \Sigma/T$ but numerical solutions are readily computed using standard optimization software. Figure 2.2 conveys the intuition behind the robust solution to a benchmark tracking constrained expected return maximization problem with two risky assets. The figure reflects the hesitance to invest if the estimated expected returns are not sufficiently different from zero. However, unlike a variance constraint, the benchmark tracking constraint may prohibit passive portfolios. In that case, the optimal solution will lie on the boundary of the feasible set. Section 2.4 reports an empirical study of a portfolio choice problem with 25 assets.

Figure 2.2: Optimal Benchmark tracking portfolios



Notes: The figure describes the optimal robust portfolio allocations (w_1, w_2) for a benchmark constrained portfolio choice problem (52) with two risky assets and an equally weighted benchmark. The axes correspond to the investments in the two assets. The contour plots circumscribe the set of robust feasible portfolios for various values of maximal tracking error τ. The solid line is the asymptote $\Omega^{-1}\mu$. The solid curve depicts the optimal investment as a function of the maximal tracking error for alternative values for θ. If the robust expected portfolio return is negative for all portfolios in the feasible set (due to large θ), the optimal solution tends to the zero vector to limit the expected loss (curve below the benchmark). Otherwise, the optimal portfolio choice departs from the benchmark towards the portfolio which maximizes the robust return $\alpha\Omega^{-1}\mu$ (upper part of the curve).

2.3 Robustness to unknown variance

The most general version of the mean-variance problem that we consider has uncertain mean and variance. The robust mean–variance problem is[3]

$$\max_{w} \min_{(\mu,\Sigma)\in\mathcal{U}\times\mathcal{S}} w'\mu - \frac{1}{2}\gamma w'\Sigma w. \tag{54}$$

[3]As the estimators for μ and Σ are independent, we consider the uncertainty set $\mathcal{U}\times\mathcal{S}$. This Carthesian product may be slightly too conservative though. We also omit the second order effect of uncertainty in the covariance matrix to the uncertainty set for the expected returns.

Let the uncertainty set of covariance matrices $\mathcal{S}$ consist of the covariance matrices with bounded distance, measured by d, from the estimated covariance matrix (e.g. the posterior covariance matrix) $\hat{\Sigma}$,

$$\mathcal{S} = \{\Sigma | \ \Sigma = \Sigma', \ d(\Sigma, \hat{\Sigma}) \leq \zeta\}. \tag{55}$$

An effective parametrization of the uncertainty set $\mathcal{S}$ should meticulously describe uncertainty but also result in a tractable optimization problem. Goldfarb and Iyengar (2003) set $\hat{\Sigma}$ to the maximum likelihood estimator of the covariance matrix and consider the distance function

$$d(\Sigma, \hat{\Sigma}) = \lambda_{\max}\left(\hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2} - I\right) \tag{56}$$

where $\lambda_{\max}(A)$ denotes the largest *absolute* eigenvalue of the matrix A. They use a Bayesian setup to quantify the maximum distance ζ. Starting from a non-informative conjugate prior distribution for Σ, the posterior distribution given the data is an inverse Wishart distribution with $T-1$ degrees of freedom and argument $\hat{\Sigma}$. This implies that the N eigenvalues λ of $\hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2}$ are independent and identically gamma distributed variables with $(T-1)/2$ degrees of freedom an scale parameter value $2/(T-1)$. Let F denote the cumulative gamma distribution with these parameters.

A bound on the maximum distance, in this case the maximal absolute eigenvalue, induces bounds on these gamma distributed variables:

$$\begin{aligned} P(d(\Sigma, \hat{\Sigma}) \leq \zeta) &= P(|\lambda_{\max} + 1| \leq \zeta) \\ &= \textstyle\prod_{i=1}^{N} P(1 - \zeta \leq \lambda_i \leq 1 + \zeta) \\ &= \left(F(1+\zeta) - F(1-\zeta)\right)^N. \end{aligned} \tag{57}$$

To associate the uncertainty set with a desired confidence level (cumulative posterior probability) p, the parameter ζ must be set equal to the unique solution of

$$\left(F(1+\zeta) - F(1-\zeta)\right)^N = p$$

This parametrization of $\mathcal{S}$ does not lead to the highest posterior density region, yet it enables a convenient solution to the inner minimization problem under a non-informative prior for Σ. Goldfarb and Iyengar (2003, lemma 3) show that for $\zeta < 1$ this form of uncertainty merely augments the covariance matrix proportionally with a factor $1/(1-\zeta)$ which is an increasing function of the desired confidence level.

The robust variance constrained feasible set will be

$$\mathcal{W}_{\sigma,R} = \{w : \ w'\hat{\Sigma}w \leq (1-\zeta)\sigma^2\}.$$

The robust solution for the special case $\Omega = \hat{\Sigma}/T$ is

$$w_{\sigma,R} = \sqrt{1-\zeta}\, w_{\sigma,r}. \tag{58}$$

The optimal solution to the expected utility version is

$$w_R = (1 - \zeta) w_r. \tag{59}$$

In this setting, robustness proportionally reduces the investment in the risky assets but employs the same preliminary test as (39). The reduction is sensitive to the number of observations. For example if $N = 25$ and the desired confidence level is 95%, then $\zeta = 0.44$ if we have 120 observations but rises to 0.91 if we have only 36 observations.

Alternatively if we include the second order effect of uncertainty in the covariance matrix to the uncertainty set $\mathcal{U}$, i.e. $\Omega = (1-\zeta)^{-1}\hat{\Sigma}/T$, the preliminary test changes to $\sqrt{1-\zeta}\hat{\rho} \geq \theta/\sqrt{T}$ and the reduction of the robust portfolio w_R increases to $(1 - \frac{\theta}{\sqrt{1-\zeta}\rho\sqrt{T}})$.

An empirical analysis of the robust portfolios is reported in section 2.4.

Benchmark tracking constrained portfolio choice

Under the same assumptions as in the mean variance framework, uncertainty in the variance for the benchmark tracking constrained problem can be robustly handled by scaling the covariance matrix. In this case the tracking error constraint becomes

$$(w - \tilde{w})' \left((1-\eta)^{-1}\Sigma\right)(w - \tilde{w}) + \left(\|\hat{\mu}'(w - \tilde{w})\| + \|\theta A(w - \tilde{w})\|\right)^2 \leq \tau^2. \tag{60}$$

2.4 Empirical results

For the expected utility maximization problem, the performance is measured by the ex-post realized utility,

$$Q_0(w) = \mu_0' w - \frac{1}{2}\gamma w' \Sigma_0 w \tag{61}$$

where μ_0 and Σ_0 denote the true mean and covariance matrix. We do not use the Sharpe ratio to measure performance as it does not take leverage into account and therefore is not an adequate measure to evaluate mean variance in the presence of uncertainty.

An actual portfolio w is a function of the sample data, the robustness preference and possibly some prior belief about the return model. The utility has a distribution depending on the distribution of the sample moments. We are interested in comparing the average performance of the mean-variance portfolio with the performance of its robust counterparts. For the robust portfolio the realized utility is a non-linear function of the sample statistics, for which we do not have a closed form expression for the density or the moments. We therefore compute these moments by simulation. The simulation is based on a bootstrap experiment with the Fama & French 25 portfolios dataset[4] consisting of 480 monthly observations from July 1963 to December 2002 on $N = 25$ (value weighted) portfolios independently sorted in size and book-to-market quintiles. For the main bootstrap experiment we use $T = 60$ observations to estimate the sample mean and covariance matrix which serve as estimates for μ and Σ. To obtain a reliable idea of portfolio performance we consider the average ex-post realized performance over $K = 10000$ bootstrap samples. We use k to refer to a particular bootstrap sample. We use w_0 to denote the true optimal mean-variance portfolio which is the optimal mean variance portfolio based on the mean and variance (μ_0, Σ_0) estimated on the entire dataset. We use w_γ^k to denote the mean-variance portfolio based on the sample moments μ^k and Σ^k of bootstrap sample k and w_r^k and w_R^k denote the mean- and mean-variance robust portfolios based on the sample moments of bootstrap sample k and the associated uncertainty sets $\mathcal{U}_k$ with $\Omega = \Sigma_k/T$ and $\mathcal{S}_k$ with $\hat{\Sigma} = \hat{\Sigma}_k$.

We denote the ex-ante expected portfolio return according to model $j = \{\gamma, r, R\}$ by r_j^k and the standard deviation by s_j^k. For example $r_\gamma^k = w_\gamma^{k\prime}\mu^k$ and $r_r^k = \min_{\mu\in\mathcal{U}^k} w_j^{k\prime}\mu$. Moreover, we define the quantities,

$$
\begin{array}{lll}
\text{ex-ante} & & \\
\quad\text{(robust) mean variance utility} & Q_j^k & = r_j^k - \frac{1}{2}\gamma s_j^{k\,2} \\
\quad\text{(robust) Sharpe ratio} & Sh_j^k & = r_j^k/s_j^k \\
\text{ex-post} & & \\
\quad\text{expected excess return} & r_0(w_j^k) & = \mu_0' w_j^k \\
\quad\text{variance} & s_0^2(w_j^k) & = w_j^{k\prime}\Sigma_0 w_j^k \\
\quad\text{mean-variance utility} & Q_0(w_j^k) & = r_0(w_j^k) - \frac{1}{2}\gamma s_0^2(w_j^k) \\
\quad\text{Sharpe ratio} & Sh_0(w_j^k) & = r_0(w_j^k)/s_0(w_j^k) \\
\quad\text{expected loss} & L_0(w_j^k) & = Q_0(w_0) - Q_0(w_j^k) \\
\text{portfolio statistics} & & \\
\quad\text{sum} & & = \iota' w_j^k \\
\quad\text{norm} & & = \|w_j^k\|. \\
\quad\text{cross-sectional st.dev.} & & = \sqrt{\|w_j^k\|^2/N - (\iota' w_j^k/N)^2}.
\end{array}
$$

Each iteration k of the bootstrap experiment performs the following tasks,

1. Randomly generate T observations from the Fama & French data.
2. Estimate the mean μ^k and covariance matrix Σ^k. Ω is set to Σ^k/T and the uncertainty in Σ^k is characterized by (55).
3. Compute the portfolios w_γ^k, w_r^k, w_R^k.
4. Compute the ex-ante performance measures r_j^k, s_j^k, Q_j^k, Sh_j^k for all portfolios j.

---

[4]Available from the homepage of Professor K. French at url `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

5. Compute the ex-post performance measures r_0, Q_0 and L_0 for all portfolios j.

6. Save the results.

We set the risk aversion parameter $\gamma = 5$. Alternative values of γ affect the total risky investment but do not affect the portfolio composition or, for robust portfolio choice, the decision whether or not to invest. Therefore the results for alternative values of γ will be scaled versions of the presented results. We choose the preference for robustness θ such that $\mathcal{U}$ is, under normality, a 95% confidence set for μ. When the variance is unknown, $(\mu - \hat{\mu})'\Omega(\mu - \hat{\mu})$ is Hotelling $T^2(N, T)$ distributed and $\theta^2 = T^2_{inv}(0,95, N, T-1) = \frac{N(T-1)}{T-N}F_{inv}(0.95, N, T-N)$, hence for $N = 25$ and $T = 60$, $\theta = 8.8$. The preference for robustness regarding the covariance estimates ζ is set such that, under the assumptions in section 2.3, $\mathcal{S}$ forms a 95% confidence set for Σ.

Table 2.1 summarizes the results for the simulation experiments. It reports averages and standard deviations of the above quantities, e.g.

$$\text{average ex-ante expected mean-variance performance} = \frac{1}{K}\sum_{k=1}^{K} Q_j^k(w_j^k). \qquad (62)$$

The keyword 'active' refers to the fraction of the bootstrap samples for which the corresponding portfolio is active, i.e. has a positive portfolio norm. $P(Q_0(w_j) \geq Q_j)$ and $P(Sh_0(w_j) \geq Sh_j)$ denote the fraction of bootstrap samples for which the ex-post mean-variance performance and Sharpe ratio respectively are at least their ex-ante expected values. Moreover $Q_{5\%}(w_j)$ denotes the lower fifth percentile of mean-variance performance in the experiment. These quantities serve as measures for robustness of the expected performance. We also report the portfolio sum and cross-sectional standard deviation which indicate leverage and the extent to which the portfolio exploits cross-sectional asset differences.

2.4.1 Results for mean-variance portfolio choice

Table 2.1, supplemented with figure 2.3, presents the results of the bootstrap experiment on the original Fama & French dataset. The null model estimated on the Fama & French dataset over the period July 1963 to December 2002 has a mean-variance performance of $Q_0(w_0) = 2.3\%$, a Sharpe ratio of 0.48, a portfolios sum of 1.64 and a cross sectional standard deviation of portfolio weights equal to 1.23.

The bias in the ex-ante mean variance performance and Sharpe ratio for the mean-variance portfolio w_γ is considerable and even larger than $\frac{1}{2\gamma}N/T$ (see (25)), which could be expected for normal distributed returns with a known covariance matrix. The robust portfolio has smaller ex-ante expected performance than the naive portfolio. This is caused by the reduction in risky investment, i.e. robustness reduces the Sharpe ratio by $\theta/\sqrt{T}$, and a preliminary test, i.e. a robust investor remains passive if the estimated

Table 2.1: Bootstrap performance of mean-variance portfolios

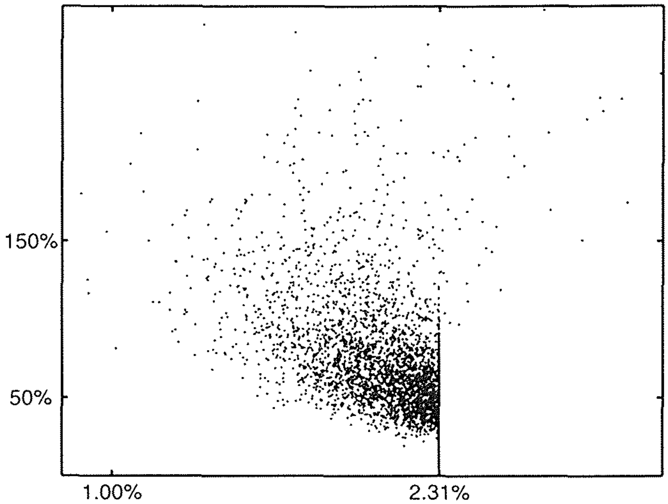
		mean robust	mean- and variance robust
	$\hat{w}_\gamma$	$\hat{w}_r$	$\hat{w}_R$
expected performance (ex-ante)			
expected return (r)	24.75	0.24	0.06
	(9.42)	(0.88)	(0.23)
standard deviation (s)	21.88	0.89	0.45
	(4.05)	(2.00)	(1.02)
mean-variance utility (Q)	12.38	0.12	0.03
	(4.71)	(0.44)	(0.11)
Sharpe ratio when active (Sh)	1.09	0.15	0.08
	(0.20)	(0.13)	(0.07)
performance under null (ex-post)			
expected return (r)	9.53	0.47	0.12
	(4.45)	(1.21)	(0.32)
standard deviation (s)	43.20	2.06	0.54
	(14.27)	(5.00)	(1.30)
mean-variance utility (Q)	-42.22	-0.26	0.07
	(35.50)	(2.52)	(0.18)
Sharpe ratio when active (Sh)	0.22	0.23	0.23
	(0.07)	(0.07)	(0.07)
expected loss (L)	44.53	2.57	2.23
	(35.50)	(2.52)	(0.18)
ex-ante to ex-post			
$P(Sh_0(w_j) \geq Sh_j)$	0.0	92.5	97.8
$P(Q_0(w_j) \geq Q_j)$	0.0	89.4	95.7
portfolio			
active	1.00	0.29	0.29
sum	3.23	0.16	0.04
cross-sectional st.dev.	6.17	0.29	0.08
$Q_{5\%}$	-0.10	0.00	0.00

Notes: The table reports the results of the bootstrap experiment as described in section 2.4 on the Fama & French 25 portfolios dataset over the period July 1963 to December 2002. Risk aversion parameter $\gamma = 5$, sample size $T = 60$ and preference for robustness $\theta = 8.9$ (95% confidence set). Ex-post performance is evaluated on the null model featuring a mean-variance performance of 2.3%, a Sharpe ratio of 0.48, a portfolio cross-sectional standard deviation of 1.23 and portfolio sum equal to 1.64. All quantities are averages or standard deviations (in brackets) resulting from $K = 10.000$ bootstrap samples and are reported in percentages, except for the Sharpe ratios and portfolios characteristics.

Sharpe ratio is not significantly different from zero ($\hat{\rho} > \theta/\sqrt{T}$). Although these robust performance measures seem to portray a pessimistic view on investment opportunities, they prove to be quite adequate considering the ex-post performance.

The participation of the robust investor in the risky market is small; only 29% of the bootstraps induce a significant Sharpe ratio and leads to active portfolios. Moreover, the robust risky investment is small and, unlike the the naive portfolio, robust portfolios are not leveraged.

Figure 2.3: Loss distribution



Notes: The figure presents a scatter plot of the ex-post losses in mean-variance performance per month ($Q_0(w_0) - Q_0(w^k)$) for a robust investor (horizontal axis) and a naive investor (vertical axis). Each point presents the losses relating to a sample of the bootstrap experiment reported in table 2.1.

The robust approach beats the naive approach in terms of robustness *and* ex-post expected performance. Indeed the expected loss for the robust investor is only a fraction of its naive competitor. We see that a robust approach to uncertainty in the expected returns leads to the largest improvement. The added value of a robust approach to uncertainty in the covariance matrix is small.

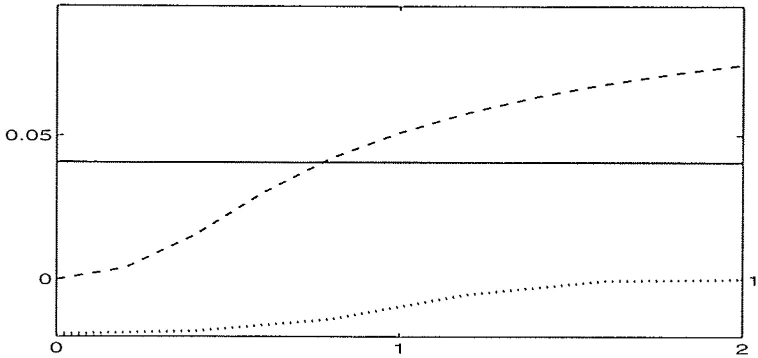
In terms of robustness, the robust portfolio yields (ex-post) the anticipated (ex-ante) performance in approximately 95% of the bootstraps and reports a nonnegative 5th percentile. The naive approach ex-ante overestimates performance and has a negative 5th percentile.

The difference between the naive and robust portfolio depends on the number of observations T, the investor's preference for robustness θ and an estimate of the economic factor: the Sharpe ratio.

Ideally a test covers every plausible economy, in particular its featured expected returns and (co)variances of the returns. We therefore repeat the experiment under alternative values of the true parameters. In particular we will consider alternative datasets with different Sharpe ratios, which is as (35) shows the main economic determinant for the difference in performance between naive and robust portfolios.

In figure 2.4 we vary the Sharpe ratio by scaling the expected return vector μ_0 of the empirical distribution such that the empirical distribution $y_t = \mu_0 + \varepsilon_t$ changes to $\tilde{y}_t =$

Figure 2.4: Sensitivity to underlying Sharpe ratio



Notes: The figure plots the expected loss in mean-variance performance per month (left scale) as a function of the underlying monthly Sharpe ratio for a mean-variance investor (solid line) and a robust mean-variance investor (dashed line). Moreover the dotted line plots the fraction of bootstrap samples (right scale), for which robust investment implies active participation in the risky assets, as a function of the underlying Sharpe ratio. This underlying Sharpe ratio is varied by scaling the mean vector of the dataset; otherwise the results are based on the framework for the bootstrap presented in section 2.4.

$\alpha\mu_0+\varepsilon_t$, and we plot the expected loss for alternative Sharpe ratios. The robust approach is not universally best; for Sharpe ratios larger than 0.77, the naive portfolio is superior to the robust portfolio in terms of expected returns. For increasing values of the Sharpe ratio, the pre-test in the robust approach becomes redundant and the robust portfolio is always active. In this case the robust portfolio is a scaled version of the naive portfolio and has expected loss

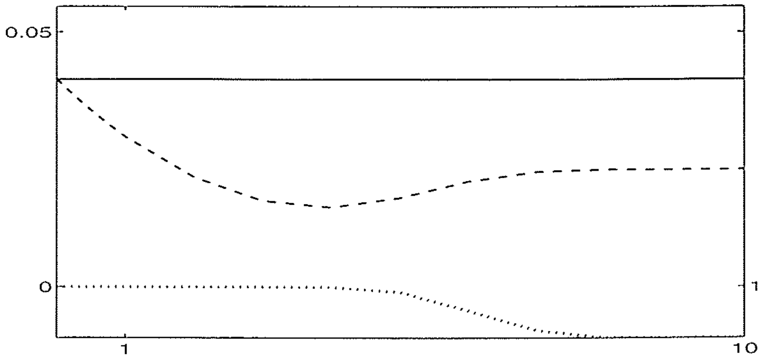
$$2\gamma \mathsf{E}[L] = 2\gamma \mathsf{E}\left[Q(w) - Q\left(\left(1 - \frac{\theta}{\hat{\rho}\sqrt{T}}\right) w_\gamma\right)\right] = \frac{N+\theta^2}{T}, \tag{63}$$

which is higher than the expected loss of the naive portfolio N/T. This is the price of a robust approach: the worst case performance over an uncertainty set of parameters is maximized to the detriment of expected loss for large Sharpe ratios.

Figure 2.5 shows the effect of alternative values for the investor's preference for robustness on expected loss. In terms of minimizing expected loss, a confidence level of less than 1% ($\theta \approx 4$) is optimal and in statistical decision theory, this confidence level would be apt for a preliminary test. In the robust approach, the investor's preference for robustness reflects the investors attitude towards uncertainty and is fixed, typically at values that correspond to a confidence level of more than 90%.

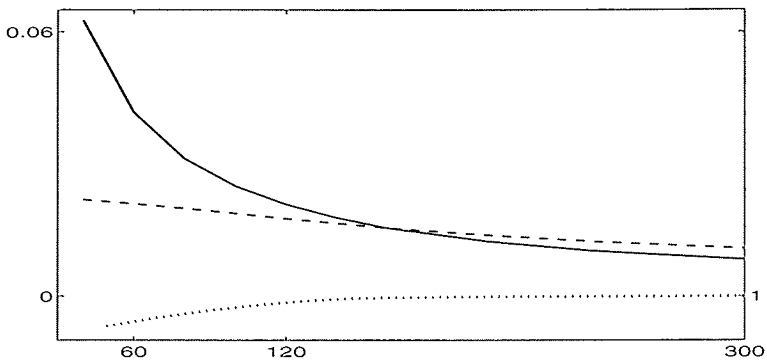
Figure 2.6 reports the expected loss as a function of uncertainty in the estimators. The effect of changes in uncertainty, measured by the number of observations, is ambiguous: in the preliminary test $\hat{\rho} \geq \theta/\sqrt{T}$ both sides of the inequality are affected by a change

Figure 2.5: Sensitivity to preference for robustness



Notes: The figure plots the expected loss in mean-variance performance per month ($Q_0(w_0) - Q_0(w^k)$) (left scale) as a function of the preference for robustness for a mean-variance investor (solid line) and a robust mean-variance investor (dashed line). Moreover the dotted line plots the fraction of bootstrap samples (right scale) for which robust investment implies active participation in the risky assets, as a function of the preference for robustness. This preference for robustness is varied by changes to θ, for example $\theta = 5.5$ and $\theta = 9$ correspond to approximately 50% and 95% confidence levels respectively. Otherwise the results are based on the framework for the bootstrap presented in section 2.4.

Figure 2.6: Sensitivity to uncertainty



Notes: The figure plots the expected loss in mean-variance performance per month ($Q_0(w_0) - Q_0(w^k)$) left scale) as a function of the extent of estimation uncertainty for a mean-variance investor (solid line) and a robust mean-variance investor (dashed line). Moreover the dotted line plots the fraction of bootstrap samples (right scale), for which robust investment implies active participation in the risky assets, as a function of the extent of estimation uncertainty. This extent of estimation uncertainty is varied by changes to the number of observations T; otherwise the results are based on the framework for the bootstrap presented in section 2.4.

in the number of observations. The right-hand side changes proportional to $1/\sqrt{T}$. On the other hand, an increase in the number of observations decreases the bias in $\hat{\rho}$ (see (24)), however this change is less than proportional to $1/\sqrt{T}$. Consequently the robust portfolio converges to the naive portfolio as the number of observations increases.

Note that a bootstrap on empirical data also tests the flexibility of the approaches to cope with small model misspecifications, in this case deviations from the normal model. Deviations from normality may render ellipsoidal uncertainty sets as inappropriate[5]. The effect to robustness of deviations from normality is limited in our experiments: the uncertainty set which is based on normality, leads to a robustness around 90%.

2.4.2 Empirical simulation

Table 2.2: Empirical simulation of mean-variance portfolios

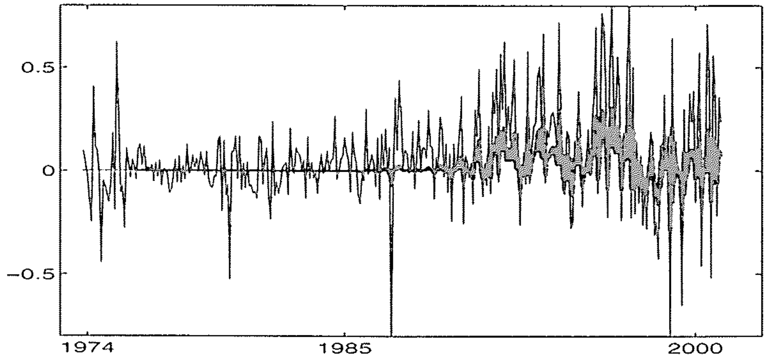
	$\hat{w}_\gamma$	$\hat{w}_r$	$\hat{w}_R$
expected performance (ex-ante)			
expected return (r)	13.66	1.29	0.65
standard deviation (s)	15.74	3.38	2.39
mean variance (Q)	6.83	0.65	0.32
Sharpe ratio (Sh)	0.79	0.17	0.12
empirical performance (ex-post)			
expected return (r)	7.36	2.23	1.12
standard deviation (s)	23.66	7.53	3.76
Sharpe ratio (Sh)	0.31	0.30	0.30
mean variance (Q)	-6.64	0.82	0.76
portfolio			
active	1.00	0.55	0.55
sum	2.45	0.77	0.38
cross	3.11	0.73	0.37

Notes: The results report on a rolling horizon simulation as described in section 2.4.2 on the Fama & French 25 portfolios dataset over the period July 1968 to December 2002. Risk aversion parameter $\gamma = 5$, sample size $T = 120$ and preference for robustness $\theta = 8.9$ (95% confidence set). Ex-post performance is evaluated on actual out-of-sample future returns. Reported quantities are means and standard deviations of 'composite' portfolio returns y, defined by (64).

By means of bootstrapping we avoided intertemporal effects unaccounted for in the static return model. Jegadeesh and Titman (1993) report that equity returns exhibit short-term (3 to 12 months) continuation so that a momentum strategy, buying assets with the best performance over the past 3 to 12 months and short selling assets with the

[5] An ellipsoidal set may still catch a 95% confidence interval if we use a non-parametric estimation like the Chebychev inequality. The disadvantage is that this estimation is not necessarily tight and therefore might lead to unnecessary conservatism.

Figure 2.7: Empirical performance



Notes: The figure shows the monthly composite portfolio returns (64) over the period 1974 to 2001 corresponding to the empirical simulation reported in table 2.2. The graphs correspond to a mean-variance investor (solid line) and a robust mean-variance investor (grey line).

worst prior performance, generates an excess return over the next few months. On the other hand, Fama and French (1988) find that 3- to 5-year equity returns are negatively serially correlated and stock prices have a tendency to revert to their trend lines over longer horizons.

In this section we conduct an historical simulation that, unlike the bootstrap experiment, leaves the time-order of the historical returns intact. The historical simulation adapts to structural breaks and long term trends in the investment opportunity set. For each month, we estimate the portfolio on past observations, and evaluate the portfolio on future, out-of-sample performance. We avoid the momentum and mean reversion effects by choosing an estimation horizon ($T = 120$) that exceeds the cycli of these effects.

The historical simulation is conducted over the period July 1974 to December 2001 with settings similar to those for the bootstrap. For each date t (total 328), the optimal portfolio w_j^t is estimated on the last ten years ($T = 120$) observations. To avoid overlap in our experiments, yet achieve significance of results, we compute the composite portfolio returns

$$y_j^t = \frac{1}{12}(r^t)'\left(w_j^{t-11} + w_j^{t-10} + \ldots + w_j^t\right), \tag{64}$$

where r^t denotes the (excess) portfolio return at date t. In other words, all portfolios are held for one year.

The results in table 2.2 show that the robust approach beats the naive approach. The robust portfolio remains passive until 1986 (figure 2.7) and reports a positive average performance over the next period.

Although the practical value of this experiment is small as no investor would actually use a static return model without any priors on the historical sample with many return dynamics, it shows that a robust portfolio, in this case to guard against estimation uncertainty, may also improve reliability under other sources of uncertainty, such as model misspecifications.

2.4.3 Benchmark tracking constrained portfolio choice

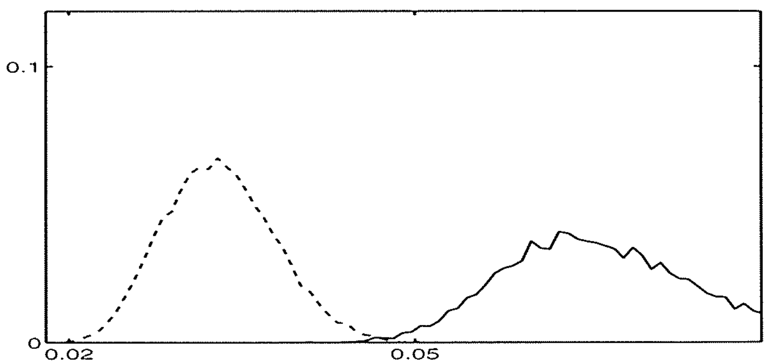
We perform similar tests for the benchmark tracking portfolios w_τ, w_b and w_B. The ex-ante tracking error associated with the naive approach follows from the left-hand side of (20) and the bootstrap sample moments μ^k and Σ^k. The tracking errors τ_b^k and τ_B^k associated with mean-robust and mean-variance robust approaches respectively follow from the left-hand sides of the inequalities in (46) and (60). Additional to the portfolio return, we will report

$$\begin{array}{llcl} \text{ex-ante} & & & \\ \quad\text{tracking error} & \tau_j^k & = & \tau_j(w_j^k) \\ \text{ex-post} & & & \\ \quad\text{tracking error} & \tau_0(w_j^k) & = & \sqrt{(w-\tilde{w})'(\Sigma_0+\mu_0\mu_0')(w-\tilde{w})}. \end{array}$$

Moreover $P(\tau_0(w_j) \leq 5\%)$ denotes the fraction of bootstrap samples for which the ex-post tracking error does not exceed the maximum allowed tracking error and serves as a measure for robustness. As the benchmark tracking limit is imposed as a hard constraint and a portfolio that violates this constraint ex-post is obsolete, this measure also reports the fraction of adequate portfolios. Moreover, $\tau_{5\%}$ denotes the upper 5th percentile of tracking error in the experiment.

For the rest we keep the experiment settings of the previous section and set the benchmark to the equally weighted portfolio.

Figure 2.8: Tracking error distribution



Notes: The figure presents the distribution of the tracking error for the bootstrap experiment reported in table 2.3. The horizontal axis measures the tracking error τ of the portfolio. The solid curve corresponds to a naive mean-variance investor, the dashed curved corresponds to distribution a robust mean-variance investor.

2.4.4 Results for benchmark tracking constrained portfolio choice

Table 2.3: Bootstrap experiment

	$\hat{w}_\gamma$	$\hat{w}_r$	$\hat{w}_R$
expected performance (ex-ante)			
expected return (r)	4.28	-1.65	-2.65
	(0.75)	(0.99)	(1.05)
standard deviation (s)	0.42	0.05	0.43
	(0.08)	(0.04)	(0.20)
tracking error τ	5.00	5.00	5.00
	(0.00)	(0.00)	(0.00)
performance under null (ex-post)			
expected return (r)	2.07	0.50	0.73
	(0.46)	(0.14)	(0.16)
standard deviation (s)	8.67	2.39	3.59
	(0.99)	(0.52)	(0.44)
tracking error τ	6.68	3.31	2.45
	(0.91)	(0.48)	(0.33)
ex-ante to ex-post			
$P(\tau_0(w_j) \leq \tau_j)$	1.02	100.0	100.0
portfolio			
active	1.00	1.00	1.00
sum of weights	1.48	0.46	0.72
cross-sectional st.dev.	0.93	0.17	0.22
$\tau_{5\%}$	8.31	4.13	3.03

Notes: The table reports on a bootstrap experiment as described in section 2.4 on the Fama & French 25 portfolios dataset over the period July 1963 to December 2002. Maximal tracking error from the equally weighted benchmark $\tau = 5\%$, sample size $T = 60$ and preference for robustness $\theta = 8.9$ (95% confidence set). Ex-post performance is evaluated on the null model featuring expected return 2.81%, a tracking error equal to 5%, a portfolio cross-sectional standard deviation of 0.58 and portfolio sum 1.76. All quantities are averages and standard deviations (in brackets) resulting from $K = 10.000$ bootstrap samples and are reported in percentages, except for the portfolios characteristics.

Table 2.3 and figure 2.8 show the results of the bootstrap experiment.

Ex-ante, the naive and robust investors differ fundamentally in their evaluation of the investment opportunity set. The naive investor believes in positive expected returns and therefore invests, more than the benchmark, in risky assets. The maximal investment in risky assets is restricted by the benchmark constraint, which is tight for the optimal naive portfolio.

The robust investor does not believe that significant positive expected returns can be attained for portfolios which satisfy the benchmark constraint. The robust expected portfolio return is naturally smaller than the expected return of the portfolio. Moreover a robust evaluation of the constraint reduces the feasible set. Therefore the robust investor tends to zero investment in risky assets, but is restricted by the tracking constraint which forces the robust portfolio to be 'active' (see table (2.3)) in the risky market. The (average) risky investment is smaller than the risky investment of the bootstrap: 0.46 for the mean-robust portfolio and 0.72 for the mean- and variance robust portfolio. The mean- and variance robust portfolio has larger risky investment because a robust approach to uncertainty in the covariance matrix leads to a feasible set which is more tightly centered around benchmark constraint with risky investment equal to 1. The robust tracking error for both robust portfolios is equal to 5% and consequently the tracking error constraint is tight.

Ex-post the average performance for a naive and robust investor change comparatively. On the one hand, the naive investor still obtains the highest expected performance, though less than expected ex-ante. On the other hand, the naive investor tramples on the benchmark constraint (see figure 2.3). The constraint is satisfied only in 1% of the bootstrap samples. The robust investor hardly ever violates the benchmark constraint.

As we impose the benchmark constraint as a 'hard' constraint, the naive approach actually has deplorable performance. We therefore conclude that the robust approach with small positive expected return and reliable tracking performance is a better portfolio choice. The mean- and variance robust approach also has better return (0.72%) than the benchmark portfolio with 0.65% expected return.

2.5 Conclusions

Empirical results show that a robust approach enables a considerable improvement of reliability at the cost, if any, of a small decrease in expected performance. A robust investor participates actively in the risky market only if expected performance is beyond some threshold value and in the empirical study we find that the robust investor often remains passive. For large uncertainty, this improves (error-maximizing) mean-variance investment considerably. Yet, a passive strategy need not be typical for a robust approach, but may ensue from the chosen return model that leads to large uncertainty in the parameters.

An investor who uses a more parsimonious model which leads to less estimation uncertainty and consequently smaller uncertainty sets, may be robust and fully active in the risky market. This connects the participation in the risky market to the investor's quality as a model builder. There is a catch though: when a structured return model is misspecified, it may reduce estimation uncertainty but introduce model uncertainty. In chapter 3, we consider the effects of model misspecification.

As already noted in section 2.4.2, a static return model is not adequate to support robust portfolio choice over time. In chapter 4 we will consider dynamic return models and a robust approach to multi-period portfolio choice in the presence of estimation uncertainty.

2.6 Supplementary Notes

The robust approach adopts preliminary tests and shrinkage to develop a type of minimax decision rule. Minimax decision rules have been studied before in *statistical decision theory* to produce estimators that minimize maximum expected loss. In this context the robust portfolio can be interpreted as a positive Stein-rule estimator that combines pretesting and shrinkage to a minimax estimator.

Applied to portfolio choice, a positive Stein-rule solves

$$\min_{w} \max_{\mu} L(\mu, w) \tag{65}$$

and is equal to (Anderson (2003, p.97)),

$$\begin{aligned} w_{mm} &= \left(1 - \frac{N-2}{T\hat{\rho}^2}\right) w_\gamma \qquad \text{if} \quad N-2 \leq T\hat{\rho}^2 \\ &= \quad 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{66}$$

The decisions in (66) and (35) are equal if $\theta = \frac{N-2}{\sqrt{T}\hat{\rho}}$. In our analysis the preference for robustness is exogenously given. Robust optimization is a behavioral assumption. The statistical development of Stein-rule estimators is motivated by reducing the expected loss of the estimator.

Another way to deal with uncertainty is a Bayesian approach. Bayesian decision makers condition, like the presented robust decision maker, on the data. The Bayesian assigns weights to alternative parameters in the form of posterior probabilities. The Bayesian approach to estimation uncertainty is to solve for an optimal *weighted* performance.

From a Bayesian point of view, estimation uncertainty about the expected returns increases the overall risk of the returns, so that the total risk in predicting future returns is the sum of uncertainty and risk. If the Bayesian chooses a prior in line with the reasoning that lead to the uncertainty set $\mathcal{U}$, the portfolio choice of a variance constrained mean-variance investor is influenced exclusively by tightening of the risk constraint:

$$\mathcal{W}_b = \{w : \ w'(\Sigma + \Omega)w \leq \sigma^2\}. \tag{67}$$

The optimal Bayesian solution for the special case $\Omega = \Sigma/T$ is

$$w_{\sigma,b} = \frac{\sigma}{\sqrt{\mu'(\Sigma+\Omega)^{-1}\mu}}(\Sigma+\Omega)^{-1}\hat{\mu} = (1+1/T)^{-1/2}w_\sigma. \tag{68}$$

The expected utility version (2) leads to a different portfolio rule mean-variance investor

$$w_{\gamma,b} = \frac{1}{\gamma}(\Sigma + \Omega)^{-1}\hat{\mu} = (1 + 1/T)^{-1}w_{\gamma}. \tag{69}$$

Both versions lead to a decrease in risky investment by scaling down the investments. This effect of uncertainty is most pronounced for the expected utility version as risky investment $w_{\gamma,b} < w_{\sigma,b}$.

Chapter 3

Model Uncertainty[1]

The axiom of correct specification
cannot be trusted in the
domain of non-experimental data.
Edward E. Leamer (1978).

In this chapter we extend the robust approach to deal with uncertainty that arises from possible return *model* misspecification.

An unstructured return model ensures a correct model specification but hampers precise parameter estimation and consequently induces estimation uncertainty which, as shown in chapter 2, may lead to poor portfolio choice. On the other hand, *a priori* structured return models may reduce estimation uncertainty but suffer model misspecification. Even so, an a priori structured model may lead to adequate portfolio decisions if the extent of model misspecification is small. We consider an investor who intents to use prior information on the model structure to reduce estimation uncertainty but who has insufficient information about which return model is most adequate for portfolio construction. The investor holds a selection of sensible and, considering the data, conceivable return model priors but cannot discriminate among these. Hence the information available to the decision maker is too vague to form an unique Bayesian prior. This is the situation that Gilboa and Schmeidler (1989) consider to motivate minimax decision rules that characterize the behavior of the robust decision maker.

Either inspired by the axioms of Gilboa and Schmeidler (1989) or by a behavioral consideration of uncertainty aversion, the robust investor chooses a portfolio that has maximal worst case performance over the set of alternative conceivable return models. The non-Bayesian decision maker behaves as an expected utility maximizer who assumes that, given her decision, the worst alternative realizes.

A truly robust portfolio choice results when we consider each return model prior *and* its associated uncertainty in the parameter estimates for establishing the worst case

[1]This chapter is based on joint work by Frank Lutgens and Peter Schotman.

performance given the investor's decision.

Figure 3.1 reflects the considerations of a robust investor with respect to model and estimation uncertainty. The solid ellipsoid corresponds to an investor who adopts an unstructured return model which we define as the return model that follows from an uninformative prior (under the maintained hypothesis that the parameters are constant). In this case only estimation uncertainty is an issue. In line with the methods described in chapter 2, the investor constructs an uncertainty set of plausible parameter values, in this case the expected returns, which she trusts to contain the true parameter value. The contours of the uncertainty set with all plausible expected returns form the solid ellipsoid. A robust decision maker who uses an unstructured return model evaluates the worst case performance on this relatively large uncertainty set of parameters.

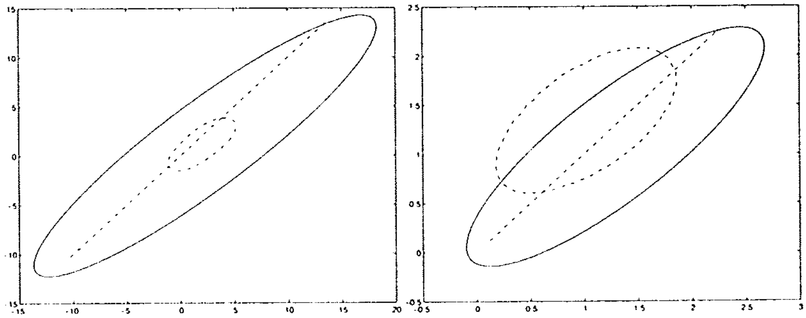
On the other hand the investor could consider informative return model priors. Figure 3.1 depicts two (dogmatic) priors by their associated uncertainty set of plausible expected returns. Both prior restrictions on the return model induce less estimation uncertainty than the unstructured model and consequently lead to a smaller uncertainty set. A decision maker who is exclusively robust to estimation uncertainty could tune her decisions to the smaller uncertainty set, yet is vulnerable to possible model misspecification. For example, this is quite likely for the return model prior in figure 2 which implies equal expected returns.

A robust decision maker who uses multiple return model priors to achieve model robustness evaluates the worst case performance on the union of the uncertainty sets that follow from the alternative prior return models. The robust investor will include enough return model priors to ensure that the correct expected returns are included in the (union of) uncertainty sets. We consider a robust investor who includes a set of restricted models but excludes the unrestricted model.

Both robust decision makers, either using an unstructured return model or multiple model priors, are confident that the true expected returns are included. The actual performance of both robust approaches depends on how well the uncertainty sets manage to capture the true return but also the size of the uncertainty sets. When the uncertainty set associated with multiple priors is well able to capture the true expected return and is smaller than the uncertainty set based on the unstructured return model, the decisions based on the former uncertainty set are preferred over more conservative decisions associated with the latter (larger) uncertainty set: Both decisions guarantee a sufficient degree of confidence but the former will have larger expected performance.

In this chapter we study a robust decision maker who consults multiple experts who propose alternative return model priors to guard against model misspecification. For some stylized portfolio choice problems we are able to derive analytically the model robust portfolio. Analysis shows that the model-robust portfolio (which is designed for best worst-case performance) also performs well in terms of expected performance. For a relevant set of plausible return models the expected performance of the robust portfolio

Figure 3.1: Model and uncertainty



Notes: The figures show the uncertainty sets containing the plausible values of the expected returns for various return models: unstructured model (solid ellipse), a return model which implies equal expected returns (dashed line) and the Fama & French factor model (dashed figure). The axes measure the expected returns of assets 5 and 18 of the Fama & French 25 dataset with 25 asset portfolios. The sets are based normal distributed errors with the last 12 (left panel) and 84 (right panel) monthly observations at December 2002 and a preference for robustness that considers all expected returns within two standard deviations from its mean.

is actually superior to the optimal naive portfolio's performance.

For general portfolio choice problems and return model priors, we can compute the robust portfolio choice numerically. This enables an empirical study which is based on the Fama & French dataset with 25 asset portfolios to test the robust portfolio choice associated with multiple return model priors, among which the Fama & French and CAPM asset pricing models. We will comapare the model robust portfolio to the optimal portfolio choice associated with the unstructured return model which is, by definition, robust to model misspecification.

3.1 Robustness to model uncertainty

To enhance intuition, we consider the effects of both sources of uncertainty, viz. model and estimation uncertainty, separately. The effect of model uncertainty hinges on the question: *What happens if the model is incorrect?*.

We consider an investor with mean variance preferences over single-period portfolio choice. Her problem is to find the optimal allocation w to risky assets with expected excess returns and covariances given by the N-vector μ and $N \times N$ covariance matrix Σ respectively. The objective function is

$$Q(w) = \mu' w - \frac{1}{2}\gamma w' \Sigma w,$$

with γ measuring risk aversion. When μ and Σ are known, the optimal investment in the risky asset is

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\mu.$$

In practice, the investor does not know the true values of μ and Σ. She obtains advice from J experts who provide her with their personal estimate (μ_j, Σ_j). The investor is convinced that these estimates are the only possible values for the true parameters. She is, however, not able to make a quantitative assessment as to which of the J parameter sets is more likely. A robust investor will consider the worst case and maximizes

$$\begin{aligned} Q_R(w) &= \min_j Q_j(w) \\ &= \min_j \mu_j' w - \frac{1}{2}\gamma w' \Sigma_j w \end{aligned} \tag{1}$$

This corresponds to what Rustem, Becker and Marty (2000) refer to as a model with rival return and rival risk scenarios. The optimal portfolio maximizes $Q_R(w)$ and is characterized by theorem 1.

Theorem 1 *Consider a robust decision maker who obtains advice (μ_j, Σ_j) from J experts. Her optimal portfolio is*

$$w_R^* = \left(\sum_{j=1}^{J} \lambda_j \Sigma_j\right)^{-1} \left(\sum_j \lambda_j \mu_j\right)$$

where λ_j are constants satisfying $0 \leq \lambda_j \leq 1$ and $\sum \lambda_j = 1$.

Proof A formal representation of the robust portfolio problem is

$$\max Q_R, \tag{2a}$$

$$\text{subject to} \quad Q_j(w) - Q_R \leq 0, \qquad (j = 1, \ldots, J). \tag{2b}$$

Kuhn-Tucker conditions for the optimal portfolio are

$$1 - \sum_j \lambda_j = 0, \tag{3a}$$

$$\sum_j \lambda_j Q_j'(w) = 0, \tag{3b}$$

$$\lambda_j (Q_R - Q_j(w)) = 0, \qquad (j = 1, \ldots, J), \tag{3c}$$

$$\lambda_j \geq 0, \qquad (j = 1, \ldots, J), \tag{3d}$$

with $\lambda_j \geq 0$ a set of Lagrange multipliers and $Q_j'(w) = \mu_j - \gamma \Sigma_j w$ the first order derivative of the objective function with expert j's parameters. From (3a) and (3d) we deduce that the Lagrange multipliers must satisfy $0 \leq \lambda_j \leq 1$. From (3b) and the linearity of $Q'_j(w)$ we immediately obtain the result stated in the theorem. □

The robust investor uses a weighted average of the opinions of the experts. In that sense the optimal portfolio is similar to the portfolio that a Bayesian investor would choose, although a Bayesian a priori assigns probabilities to the advice of each of the experts, whereas the pseudo probabilities of the robust investor are derived endogenously from solving the maximization problem.

Non-zero, positive weights are associated with active constraints and in view of (2b) refer to the models, characterized by the parameter pair (μ_j, Σ_j), that induce the worst case performance for the robust optimal portfolio. The theorem provides a partial solution in the sense that the λ_j are not given explicitly. A complete analytical solution for the minimization (1) with general μ_j and Σ_j is not available, but two special cases are analytically tractable and provide us with the intuition on robust portfolio choice.

One risky asset, two experts

Consider an investor whose choice is limited to one risky asset and the risk free asset. We assume that the expected excess return on the single risky asset is positive and consequently the optimal risky investment is non-negative. When we allow for borrowing at the risk free rate, the optimal portfolio could be $w > 1$. The quadratic objective function (3.1) reduces to $Q(w) = \mu w - \frac{1}{2}\gamma\sigma^2 w^2$ where σ^2 is the variance of the risky asset and w is a scalar in this case. The optimal portfolio w^* in (3.1) reduces to $w^* = \frac{\mu}{\gamma\sigma^2}$ with objective value

$$Q(w^*) = \frac{\mu^2}{2\gamma\sigma^2} \tag{4}$$

Also observe that the objective is zero at $w = 0$ and $w = 2w^*$ and is positive for $0 < w < 2w^*$.

The investor does not know the true values of μ and σ^2. She gets advice from two experts who have different opinions: (μ_1, σ_1^2) and (μ_2, σ_2^2). The investor is convinced that these are the only possible values for the true parameters. She is, however, not able to make a quantitative assessment which of the two parameter sets is more likely.

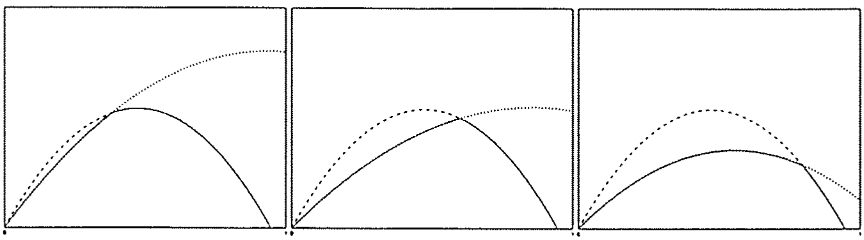
Suppose she considers a robust investment decision and maximizes the worst case utility

$$Q_R(w) = \min_{j=1,2} Q_j(w) \tag{5}$$

Before explicitly solving the robust portfolio problem, let us first distinguish different cases for (μ_j, σ_j^2). First, if $\mu_1 \geq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$, expert 1 is more optimistic than expert 2: She expects higher returns with lower risk. Consequently, the worst case $\min(Q_1(w), Q_2(w)) = Q_2(w)$ for all $w > 0$ and the optimal portfolio is the optimal portfolio according to expert 2. The opinion of expert 1 does not affect the decision at all. Conversely, if $\mu_1 \leq \mu_2$ and $\sigma_1^2 \geq \sigma_2^2$, we have the mirror image with only the advice of expert 1 being relevant.

The interesting case is when $\mu_1 > \mu_2$ and $\sigma_1^2 > \sigma_2^2$ (or both inequalities reversed). Figure 3.2 depicts the mean-variance functions $Q_j(w)$ ($j = 1, 2$). Since $Q_j(0) = 0$, both functions start at the origin. The objective function of the robust investor is the minimum of $Q_1(w)$ and $Q_2(w)$. The optimal portfolio is located at the maximum of the robust objective function.

Figure 3.2: Robust Portfolio Choice.



The figure shows the mean-variance objective functions implied by the advice of two experts: expert one (dashed line) with largest expected returns and variances and expert two (dotted line). The solid line is the minimum of the two objective functions and presents the robust objective function. The horizontal axis shows the portfolio weight of the risky asset, the vertical axis shows the value of the mean-variance objective. In the left panel the robust optimum coincides with the optimum of expert one, in the right panel with the optimum of expert two, and in the middle panel with the portfolio for which the objective values for both experts are equal.

Depending on the parameter values, three situations can occur as shown in figure 3.2. The optimum may be the maximum of $Q_1(w)$, the point where $Q_1(w)$ and $Q_2(w)$ cross or the maximum of $Q_2(w)$. Theorem 2 explicates the optimal robust decision.

Theorem 2 *Suppose $\mu_1 > \mu_2$ and $\sigma_1^2 > \sigma_2^2$. Let w_j^* be the optimal portfolio according to expert j's advice. Further define $w_{12} > 0$ as the portfolio weight for which $Q_1(w) = Q_2(w)$. A robust investor will have an optimal portfolio*

$$w_R = \lambda w_1^* + (1 - \lambda) w_2^* \tag{6}$$

with

$$\lambda = \begin{cases} 0 & \text{if} \quad \mu_2 \leq \frac{2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\mu_1, \\ \frac{2(\mu_1-\mu_2)\sigma_1^2\sigma_2^2-\mu_2\sigma_1^2}{(\mu_1\sigma_2^2-\mu_2\sigma_1^2)(\sigma_1^2-\sigma_2^2)} \in (0,1) & \text{if} \quad \frac{2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\mu_1 < \mu_2 < \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2}\mu_1, \\ 1 & \text{if} \quad \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2}\mu_1 \leq \mu_2 < \mu_1, \end{cases} \tag{7}$$

Proof With only two experts we only need to check three cases of combinations of active constraints in the Kuhn-Tucker conditions in the proof of theorem 1. We will show that these correspond to the three possible portfolios in (6). Under the assumption $\mu_1 > \mu_2$, we have $Q_1'(0) = \mu_1 > \mu_2 = Q_2'(0)$, implying that $Q_R(w) = \min(Q_1(w), Q_2(w)) = Q_2(w)$ for small w. Since both objective functions are quadratic, the difference between them is also quadratic, and they intersect at only two points: $w = 0$ and $w = w_{12}$. Therefore

$$Q_R(w) = \begin{cases} Q_2(w) & \text{if } 0 < w < w_{12}, \\ Q_1(w) & \text{if } w \geq w_{12}. \end{cases} \tag{8}$$

The maximum of $Q_R(w)$ is either at one of the interior maxima w_1^* or w_2^*, or at the intersection point w_{12}. Differentiating $Q_j(w)$, the interior maxima are easily found as

$$w_j^* = \mu_j/\gamma\sigma_j^2. \tag{9}$$

Since the expected returns are assumed to be positive, the optimal investment is always positive. We therefore do not need to consider portfolios with $w < 0$. The intersection point w_{12} follows from

$$Q_1(w) = Q_2(w) \quad \Leftrightarrow \quad (\mu_1 - \mu_2)w = \frac{1}{2}\gamma(\sigma_1^2 - \sigma_2^2)w^2,$$

so that

$$w_{12} = \frac{2(\mu_1 - \mu_2)}{\gamma(\sigma_1^2 - \sigma_2^2)}. \tag{10}$$

Let us first check if w_1^* can be a valid optimum. It is only a valid optimum if $w_1^* > w_{12}$, as otherwise $Q_R(w) = Q_2(w)$. Using the expressions for w_1^* and w_{12} this leads to the inequality

$$\frac{\mu_1}{\gamma\sigma_1^2} > \frac{2(\mu_1 - \mu_2)}{\gamma(\sigma_1^2 - \sigma_2^2)}, \tag{11}$$

which reduces to

$$\mu_2 > \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}\mu_1. \tag{12}$$

Since $\sigma_1^2 > \sigma_2^2$, we have $\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2} < 1$, and there exist pairs (μ_1, σ_1^2) and (μ_2, σ_2^2) for which the inequality holds. Moreover, if w_1^* is a valid local optimum, it is also the global optimum, since a local optimum $\tilde{w}$ of $Q_2(w)$ can only be valid if $Q_2(\tilde{w}) < Q_1(\tilde{w})$.

An analogous argument leads to the condition for w_2^* to be a valid optimum. Solving the inequality $w_2^* < w_{12}$ gives

$$\mu_2 < \frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\mu_1. \tag{13}$$

Again, since we have assumed $\sigma_1^2 > \sigma_2^2$, this optimum can occur for sufficiently small μ_2. Inequalities (12) and (13) can not hold simultaneously since

$$\frac{2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} < \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2}. \tag{14}$$

Therefore there also exists an interval

$$\frac{2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\mu_1 < \mu_2 \leq \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2}\mu_1, \tag{15}$$

for which the robust optimal portfolio is the corner solution w_{12}. □

From the graph we see that a robust portfolio is not necessarily an extremely conservative portfolio. In the middle panel of figure 3.2 the robust portfolio $w_R = w_{12}$ allocates a larger share to the risky asset than w_1^*.

Given a portfolio choice, the performance of the investment depends on the true parameters. The *loss function* expresses the utility loss that the investor incurs if she selects portfolio w relative to the optimal portfolio when the true parameters are (μ, σ^2). Formally,

$$\begin{aligned} L(\mu, \Sigma|w) &= \max_w Q(w) - Q(w) \\ &= \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu - \mu'w + \frac{1}{2}\gamma w'\Sigma w \\ &= \frac{1}{2\gamma}\left(\mu - \gamma\Sigma w\right)'\Sigma^{-1}\left(\mu - \gamma\Sigma w\right) \end{aligned} \tag{16}$$

Since we can interpret $Q(w)$ as the risk adjusted expected return of a portfolio, the loss function measures the expected return difference between the optimal portfolio and a sub-optimal portfolio, both after risk adjustment. The alternative interpretation of the loss function is the utility loss when the investor decides on portfolio w while the true parameters are (μ, σ^2).

The loss function (16) for the single-asset example simplifies to

$$L(\mu, \sigma^2|w) \quad = \quad \frac{\mu^2}{2\gamma\sigma^2}\left(1 - \frac{\gamma\sigma^2 w}{\mu}\right)^2. \tag{17}$$

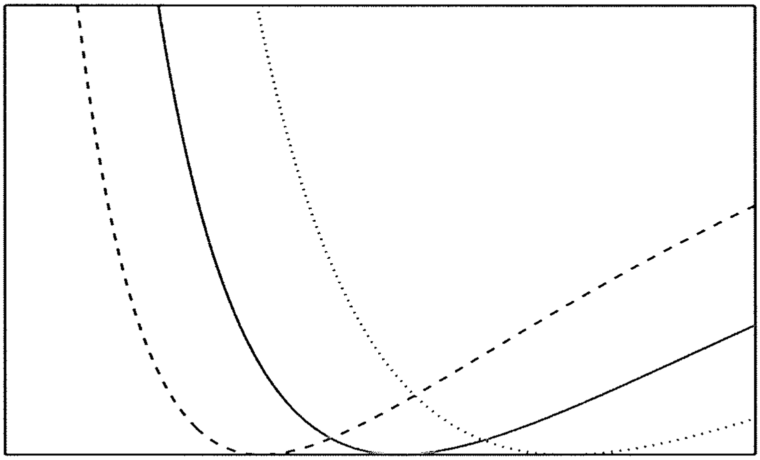
It is more convenient to consider the relative utility loss

$$\tilde{L}(\mu/\sigma^2|w) = \frac{L(\mu,\sigma^2|w)}{Q(w^*|\mu,\sigma^2)} = \left(1 - \frac{\gamma w}{\mu/\sigma^2}\right)^2, \tag{18}$$

which is solely a function of μ/σ^2.

A typical situation is depicted in figure 3.3 which shows the loss function for the portfolios w_1^*, w_2^* and $w_R = w_{12}$ for the same parameters as the upper panel of figure 3.2. None of the three portfolio rules dominates the others for all possible parameter values. For small μ/σ^2 portfolio w_1^* is best, for large μ/σ^2 the best portfolio is w_2^*, while the robust portfolio is best for intermediate parameter values. The worst robust portfolio performance is, by construction, at least as good as the worst case performance of either of its constituent models.

Figure 3.3: Expected Loss



The figure shows the relative loss functions $\tilde{L}(\mu/\sigma^2)$ (vertical axis) as a function of μ/σ^2 (horizontal axis) for three portfolios: the optimal portfolios according to the advice of experts 1 and 2 (dashed lines) for which the expected return are depicted in figure 3.2 and the robust portfolio $w_R = w_{12}$ (solid line).

The robust portfolio is *never worse* than both constituent models. Indeed, w_{12} is a strictly convex combination of w_1^* and w_2^* and the loss functions (17) and (18) are strictly convex functions which implies that the (relative) utility loss of the robust portfolio is strictly smaller than the largest loss of the two alternatives. Moreover, the robust portfolio may *outperform*, as in figure 3.3, both alternatives for a range of parameter values that fall in-between the advices of the experts. When we consider that the different experts advices follow from alternative beliefs about return models and that these different beliefs possibly outline a set of parameter values which contains the true parameter values (μ/σ^2), the robust portfolio is likely to outperform either of the portfolios based on a single advice.

Multiple risky assets with known covariance matrix

As a second special case we consider the multivariate robust problem with uncertainty in the mean excess returns μ but with a known covariance matrix Σ. When the covariance matrix is known, theorem 1 implies that the optimal robust portfolio is a convex combination of the optimal portfolios associated with the alternative (expected) return models. When the investor consults only two experts with advice (μ_1, Σ) and (μ_2, Σ), respectively, the optimal portfolio allocation is summarized in theorem 3 below.

Theorem 3 *Consider a robust portfolio optimization problem with multiple assets and two experts with different expected returns μ_1 and μ_2, but a common covariance matrix*

Σ. *Define* $\rho_{ij} = \mu_i'\Sigma^{-1}\mu_j$. *The optimal robust portfolio is*

$$w_R^* = \lambda w_1^* + (1-\lambda)w_2^* \tag{19}$$

with

$$\lambda = \begin{cases} 0 & \textit{if} \quad \rho_{22} \leq \rho_{12} \\ \frac{\rho_{22}-\rho_{12}}{(\rho_{11}-\rho_{12})+(\rho_{22}-\rho_{12})} & \textit{if} \quad \min(\rho_{11},\rho_{22}) > \rho_{12} \\ 1 & \textit{if} \quad \rho_{11} \leq \rho_{12}, \end{cases} \tag{20}$$

and $w_j^* = \frac{1}{\gamma}\Sigma^{-1}\mu_j$ *and*

Proof The robust problem with two experts can be formulated as (2a)-(2b). According to the Kuhn-Tucker optimality conditions, we need to check three possibilities: either one of the constraints is active or both constraints are active. Let us start by defining the robust criterion function as

$$Q_R(w) = \min_{j=1,2} Q_j(w) = \begin{cases} Q_1(w) & \text{if } (\mu_1-\mu_2)'w \leq 0, \\ Q_2(w) & \text{if } (\mu_1-\mu_2)'w \geq 0. \end{cases} \tag{21}$$

Next we check when w_1^* is a valid optimum. For this we need the inequality

$$(\mu_1-\mu_2)'w_1^* < 0, \tag{22}$$

which by direct substitution for w_1^* is equivalent to $\rho_{11} \leq \rho_{12}$. Analogously, the interior optimum w_2^* is valid if $\rho_{22} \leq \rho_{12}$. Since the Cauchy-Schwartz inequality implies that $\rho_{11}\rho_{22} > \rho_{12}^2$, both inequalities can not hold simultaneously. On the other hand, it is possible that both constraints are active, in which case we obtain the solution from (3b) as

$$w = \frac{1}{\gamma}\Sigma^{-1}\left(\lambda\mu_1 + (1-\lambda)\mu_2\right). \tag{23}$$

The value of the Lagrange multiplier λ then follows from the constraint $Q_1(w) = Q_2(w)$, such that

$$(\mu_1-\mu_2)'\,w = 0 \tag{24}$$

Substituting (23) into (24) gives the result of the theorem. □

The important numbers are the quadratic forms $\rho_{ij} = \mu_i'\Sigma^{-1}\mu_j$. The diagonal elements ρ_{11} and ρ_{22} are the optimal ex-ante quadratic Sharpe ratios implied by the advice of experts 1 and 2, respectively. The cross product ρ_{12} can be interpreted as the cross sectional covariance of portfolios w_1^* and w_2^*, and the ratio $\frac{\rho_{12}^2}{\rho_{11}\rho_{22}}$ as the cross sectional correlation between the implied portfolios. The lower the correlation, the more disparate the advice of the two experts, and the larger the incentive of the robust investor to combine the advice of the two experts. The larger the cross sectional variance of expert 1's portfolio, the lower the weight given to this portfolio.

The following special choice of expert beliefs gives additional insight in the implications of robust portfolio choice.

Theorem 4 *Let the advice of expert 1 be some unrestricted vector of expected returns μ_1, for example the sample means of historical returns. Let expert 2's advice be given by a projection of these same returns on a factor space with $(N \times K)$ matrix of factor loadings B,*

$$\mu_2 = B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\mu_1 \tag{25}$$

Then the optimal robust portfolio coincides with the portfolio advice of expert 2.

Proof The expected returns μ_2 are nothing but the GLS projection of μ_1 on the factor loadings B. Straightforward calculation gives

$$\begin{aligned}\rho_{12} &= \mu_1'\Sigma^{-1}\mu_2 \\ &= \mu_1'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\mu_1 \\ &= \rho_{22}\end{aligned} \tag{26}$$

Therefore, by (19), $w_R^* = w_2^*$. □

If the expert models are nested, the robust investor will follow the advice from the most restricted model. The result is very similar to the Bayesian decision process. It is a well known result in Bayesian statistics that a posterior odds comparison of nested models leads to the choice of the restricted model in case the prior on the general model is non-informative.

The loss function associated with each of the portfolios w_1^*, w_2^* and w_R^* follows from the general formula (16). Theorem 5 compares the performance of the robust investor with the performance of an investor who follows the advice of only one of the experts.

Theorem 5 *Assume $\rho_{12} < \min(\rho_{11}, \rho_{22})$. The robust portfolio w_R^* dominates the portfolios w_1^* and w_2^* for a range of true expected returns μ_0 satisfying*

$$\frac{1}{2}(\rho_{12} - \rho_{11}) < \rho_{02} - \rho_{01} < \frac{1}{2}(\rho_{22} - \rho_{12}), \tag{27}$$

where $\rho_{0j} = \mu_0'\Sigma^{-1}\mu_j$ $(j = 1, 2)$.

Proof Under the assumption the robust portfolio is a convex combination of w_1^* and w_2^*, and not one of the extremes. Substituting the portfolio weight w_R^* in the loss function (16) with true expected returns μ_0 we obtain

$$\begin{aligned}2\gamma L(\mu_0|w_R^*) &= (\mu_0 - \gamma\Sigma w_R^*)'\,\Sigma^{-1}\,(\mu_0 - \gamma\Sigma w_R^*) \\ &= (\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2)\,\Sigma^{-1}\,(\mu_0 - \lambda\mu_1 - (1-\lambda)\mu_2) \\ &= \rho_{00} + \lambda^2\rho_{11} + (1-\lambda)^2\rho_{22} \\ &\qquad - 2\lambda\rho_{01} - 2(1-\lambda)\rho_{02} + 2\lambda(1-\lambda)\rho_{12}\end{aligned} \tag{28}$$

The analogous expression for the utility loss of expert 1 is

$$2\gamma L(\mu_0|w_1^*) = \rho_{00} + \rho_{11} - 2\rho_{01} \tag{29}$$

Subtracting (29) from (28) gives

$$\begin{aligned} 2\gamma\left(L(\mu_0|w_1^*) - L(\mu_0|w_R^*)\right) = {} & (1-\lambda^2)\rho_{11} - (1-\lambda)^2\rho_{22} \\ & - 2\lambda(1-\lambda)\rho_{12} - 2(1-\lambda)\rho_{01} + 2(1-\lambda)\rho_{02} \end{aligned} \tag{30}$$

For the robust investor to have lower loss, this must be positive. Divide (30) by $1-\lambda$, require the result to be positive and solve for $\rho_{02} - \rho_{01}$,

$$\begin{aligned} 2(\rho_{02} - \rho_{01}) & > -(\lambda+1)\rho_{11} + (1-\lambda)\rho_{22} + 2\lambda\rho_{12} \\ & = -\rho_{11} + \rho_{22} - \lambda(\rho_{11} + \rho_{22} - 2\rho_{12}) \\ & = \rho_{12} - \rho_{11} \end{aligned} \tag{31}$$

The last line follows from the definition of λ. The result establishes the first inequality in the theorem.
Proof of the second inequality is analogous by comparing $L(\mu_0|w_2^*)$ and $L(\mu_0|w_R^*)$. □

Theorem 5 implies that the N-dimensional expected return space is cut by two parallel planes in between which the robust portfolio has lower expected loss than either of the two expert portfolios. By convexity considerations, the robust portfolio has better performance than the worst alternative beyond that space. Moreover, the expected return space outlined by alternative experts' advice may well contain the true parameter value. Hence the robust portfolio may also have better actual performance than either alternative portfolio.

General return model priors

We are not able to derive analytical solutions to the robust portfolio choice problem (2a)-(2b) with multiple assets and experts who propose general alternative return model priors. Instead we may compute the robust solution of the quadratically constrained optimization problem numerically (see Rustem et al. (2000)), at least if the number of alternative return model priors and consequently the number of constraints (2b) is finite.

3.2 Multiple model priors

So far we have not been explicit on how the experts arrive at their estimates (μ_j, Σ_j). When defining the loss function we ignored the relation between the reported advice (μ_j, Σ_j) and the true parameters. In this section we describe the advice of the experts. To advise an investor, the expert combines her return model prior with the sample's information. Depending on the expert's confidence in the return model prior and the strength of sample information, she obtains a posterior return model by applying Bayes rule. This posterior return model is the expert's advice to the investor.

We assume that experts have non-dogmatic return model priors so that the alternative posterior return models converge as more observations are added. Moreover one could use posterior odds to select the experts with sensible model priors, at least if the priors are proper. On the other hand, an investor may prefer to characterize uncertainty by other means than (evaluating the posterior odds of the priors on) the data. We adopt the latter reasoning and assume that the robust investor is confident that the experts' advice is sensible.

In this section we first consider the optimal portfolio associated with an unstructured model which, by definition, does not suffer model misspecification. Secondly we derive the optimal portfolios associated with alternative structured models. The model-robust portfolio will be based on these alternative structured return models.

Unstructured model

Apart from any structured and possibly misspecified model, one could consider a single expert with a non-informative prior to estimate the expected return vector and covariance matrix. Consider an expert who knows the true covariance matrix and uses a non-informative prior such that the posterior estimate of the expected return vector is the sample mean $\bar{y}$. Her advice to the investor is thus $(\bar{y}, \Sigma)$. If the investor chooses the optimal portfolio according to this advice, her loss function becomes[2]

$$L(\mu|w_{MV}) = \frac{1}{2\gamma}(\mu - \bar{y})\Sigma^{-1}(\mu - \bar{y}) \tag{32}$$

So far we always conditioned on the advice. Integrating over the sampling variation the expected loss is

$$\mathsf{E}[L(\mu|w_{MV})] = \frac{1}{2\gamma}\mathsf{E}[tr(\Sigma^{-1}(\mu - \bar{y})(\mu - \bar{y})')] = \frac{N}{2\gamma T}. \tag{33}$$

Even though the expert uses a correct model, there is still estimation error, and therefore the expected loss is not equal to zero. Note, however, that the expected loss is constant and does not depend on the true parameters (μ, Σ).

When the covariance matrix is unknown the expert uses the sample covariance matrix $\hat{\Sigma}$ as an estimate. The loss function changes to

$$L(\mu, \Sigma|w_{MV}) = \frac{1}{2\gamma}(\mu - \Sigma\hat{\Sigma}^{-1}\bar{y})'\Sigma^{-1}(\mu - \Sigma\hat{\Sigma}^{-1}\bar{y}) \tag{34}$$

[2]In case we assume that the covariance matrix is known we omit Σ as an argument in the loss function.

From classical multivariate statistics we know that under normality $\bar{y}$ and $\hat{\Sigma}$ are independent. First taking expectations over $\bar{y}$ conditional on $\hat{\Sigma}$, gives the expected loss

$$\mathsf{E}[L(\mu, \Sigma | w_{MV}) | \hat{\Sigma}] = \frac{1}{2\gamma}\left(\mu'(\Sigma^{-1} - \hat{\Sigma}^{-1})\Sigma(\Sigma^{-1} - \hat{\Sigma}^{-1})\mu + \frac{1}{T}tr(\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\Sigma)\right) \quad (35)$$

The next step would be integrating over $\hat{\Sigma}$, but this can not be done analytically. We will calculate the loss function by simulation in section 3.3 below.

Structured models

Asset pricing theories typically present a linear factor structure for expected returns. We consider experts who propose model priors with k observed factors x_{jt} as in (Pàstor and Stambaugh 2000).

A k-factor model for excess returns y_t is

$$y_t = \alpha + Bx_t + u_t, \qquad \mathsf{E}[u_t u_t'] = D \quad (36)$$

where y_t and α are N-vectors of excess and exceptional returns respectively, B is a $(N \times k)$ matrix, x_t a vector of length k, and u_t contains error term with mean zero and $(N \times N)$ covariance matrix D. The factor model is

$$x_t = \nu + e_t, \qquad \mathsf{E}[e_t e_t'] = \Psi \quad (37)$$

with ν a k-vector with factor means and e_t a vector of shocks with mean zero, uncorrelated with u_t, and covariance matrix Ψ.

A dogmatic believer of a pure factor model assumes $\alpha = 0$ and D diagonal. Expected excess returns under the factor model are

$$\mu = B\nu \quad (38)$$

and the covariance matrix

$$\Sigma = B\Psi B' + D. \quad (39)$$

However we consider non-dogmatic experts so that their posterior return models converge as more data become available. Following Pàstor and Stambaugh (2000), we consider experts with a prior on α, B, ν, D and Ψ:

$$
\begin{aligned}
D^{-1} &\sim W(df, H^{-1}) \\
\alpha \mid D &\sim N(0, \frac{\sigma_\alpha^2}{s^2} D) \\
p(B) &\propto 1 \\
p(\nu) &\propto 1 \\
p(\Psi) &\propto |\Psi|^{-(k+1)/2}
\end{aligned} \tag{40}
$$

with $H = s^2(df - N - 1)I_N$ and, following an empirical Bayes approach, the value of s^2 is set equal to the average of the diagonal elements of the sample estimate of D. $W(df, S)$ denotes the Wishart distribution with df degrees of freedom and parameter S.

Observe that the prior for D approximately contains the equivalent information of df observations and the expectation of D equals $s^2 I_N$. The strength of the expert's believe in the factor model is quantified by two coefficients: σ_α and df. A dogmatic follower of the factor model will set $\sigma_\alpha = 0$. Someone with serious doubts about the factor model will set σ_α large and df small.

To derive the posterior estimators, define the data matrices $Y = (y_1', \ldots, y_T')$, $X = (x_1', \ldots x_t')$ and $Z = (\iota_T\ X)$ with dimensions $(T \times N)$, $(T \times k)$ and $(T \times (k+1))$ respectively and the $((k+1) \times N)$ matrix $A = [\alpha\ B]'$ and denote $a = vec(A)$. Also define the statistics $\hat{A} = (Z'Z)^{-1}Z'Y$, $\hat{a} = vec(\hat{A})$, $\hat{D} = (Y - Z\hat{A})'(Y - Z\hat{A})/T$, $\hat{\nu} = X'\iota_T/T$ and $\hat{\Psi} = (X - \iota_T\hat{\nu}')'(X - \iota_T\hat{\nu}')/T$.

Pàstor and Stambaugh (2000) report the posterior

$$
\begin{aligned}
D^{-1} \mid X,\ Y &\sim W(T + df - k, (H + T\hat{D} + \hat{A}'Q\hat{A})^{-1}) \\
\alpha \mid D,\ X,\ Y &\sim N(I_N \otimes F^{-1}Z'Z)\hat{\alpha}, D \otimes F^{-1})
\end{aligned} \tag{41}
$$

with $F = diag(s^2/\sigma_\alpha^2,\ 0_k') + Z'Z$ where 0_k presents a vector with k zeros and $Q = Z'(I_T - ZF^{-1}Z')Z$ and

$$
\begin{aligned}
\Psi^{-1} \mid X,\ Y &\sim W(T - 1, \hat{\Psi}^{-1}/T) \\
\nu \mid \Psi,\ X,\ Y &\sim N(\hat{\nu}, \Psi/T)
\end{aligned} \tag{42}
$$

As B and ν are independent in the posterior, the posterior mean of the expected return vector is

$$
\tilde{\mu} = \mathsf{E}[\mu|\ X, Y] = \tilde{\alpha} + \tilde{B}\tilde{\nu}. \tag{43}
$$

where the tilde refers to the posterior expectation given the data. An analytical expression for the posterior variance

$$
\tilde{\Sigma} = \mathsf{E}[\Sigma|\ X, Y] = \mathsf{E}[B\Psi B' + D|\ X, Y] = \mathsf{E}[B\Psi B'|\ X, Y] + \tilde{D} \tag{44}
$$

is not available. We could (approximately) compute the posterior variance by Monte-Carlo simulation. We will use $\tilde{\Sigma} = \tilde{B}\tilde{\Psi}\tilde{B}' + \tilde{D}$ in our computations. This actually underestimates the posterior covariance matrix, but as the (co)variances of the elements of B (see (41)) are typically small, the underestimation is limited (about 1% in our empirical study).

A MV investor chooses the portfolio

$$w^*_{FM} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}(\tilde{\alpha} + \tilde{B}\tilde{\nu}) \tag{45}$$

If the investor chooses the optimal portfolio according to (45), her loss function becomes

$$L(\mu, \Sigma|\bar{y}) = \frac{1}{2\gamma}(\mu - \gamma\Sigma w^*_{FM})\Sigma^{-1}(\mu - \gamma\Sigma w^*_{FM}) \tag{46}$$

If α, B, Σ and Ψ are known such that $w^*_{FM} = \frac{1}{\gamma}\Sigma^{-1}B\tilde{\nu}$, we can simplify the loss function

$$\begin{aligned}
2\gamma\mathsf{E}[L(\mu, \Sigma|\bar{y})] &= \mathsf{E}[(\mu - \alpha - B\tilde{\nu})'\Sigma^{-1}(\mu - \alpha - B\tilde{\nu})] \\
&= \mathsf{E}[(\mu - \alpha - B\nu - B\bar{e})'\Sigma^{-1}(\mu - \alpha - B\nu - B\bar{e})] \\
&= \big((\mu - \alpha - B\nu)'\Sigma^{-1}(\mu - \alpha - B\nu) + \mathsf{E}[\bar{e}'B'\Sigma^{-1}B\bar{e}]\big) \\
&= \big((\mu - B\nu)'\Sigma^{-1}(\mu - B\nu) + \mathsf{E}[\mathsf{tr}(B'\Sigma^{-1}B\bar{e}\bar{e}')]\big) \\
&= \big((\mu - B\nu)'\Sigma^{-1}(\mu - B\nu) + \mathsf{tr}(B'\Sigma^{-1}B\Psi/T)\big) \\
&= (\mu - B\nu)'\Sigma^{-1}(\mu - B\nu) + \frac{1}{T}\mathsf{tr}(I + D(B\Psi B')^{-1})^{-1}
\end{aligned}$$

Observe that the second term in the last expression is smaller than k (and (33)), indicating a smaller loss through estimation errors. The additional, positive term arises from model misspecification. The loss can be both larger or smaller than in the model without misspecification. The restrictions of the factor model lead to less estimation risk, but when the factor assumption differs from the true means, model misspecification will increase expected loss.

For the empirical part of the paper we assume that the experts select their return model priors from two well known factor models: the Capital Asset Pricing Model (CAPM) and the Fama & French factor model. The CAPM developed primarily by Sharpe (1964), Lintner (1965) and Mossin (1966), employs one factor: excess market return. The factor sensitivities are referred to as asset beta's. The Fama & French model has two additional factors, firm size (SMB) and book to market (HML). The strength of the investor's belief in her favorite factor models is expressed by σ_α and df. We consider four experts by combining either factor model with two alternative degrees of believe in the corresponding factor model with $\sigma_\alpha = 0.1\%$ and $\sigma_\alpha = 4\%$ (annually for $T = 60$ observations) respectively, while we keep the prior information on to the covariance matrix low ($df = 30$).

3.3 Model robust: Empirical study

In this section we compare the empirical ex-post performance of portfolios based on (i) the unstructured return model, (ii) alternative, structured return models proposed by alternative experts and (iii) the portfolio of a robust investor who considers all alternative, structured return models proposed by the experts. The comparison is based on a bootstrap experiment on the Fama & French dataset[3] consisting of 480 monthly observations from July 1963 to December 2002 on 25 value weighted portfolios independently sorted in size and book-to-market quintiles. We refer to the sample mean and covariance matrix taken over the entire dataset as the null model (μ_0, Σ_0).

The ex-post performance of a portfolio w is given by

$$Q_0(w) = \mu_0' w - \frac{1}{2}\gamma w' \Sigma_0 w \tag{47}$$

with the risk aversion parameter[4] set to 5. We do not use the Sharpe ratio to measure performance as it does not take leverage into account and therefore is not an adequate measure to evaluate mean-variance performance in the presence of uncertainty.

An actual portfolio w is a function of the prior, sample data and, for the robust portfolio, the preference for robustness. Given a bootstrap sample of T observations from the Fama & French dataset, the portfolio w_u is based on the sample estimates, the portfolio w_c is based on the CAPM prior with $\sigma_\alpha = 0.1\%$ and the sample, w_d is based on the CAPM prior with $\sigma_\alpha = 4\%$ and the sample, w_F is based on the Fama & French prior with $\sigma_\alpha = 0.1\%$ and the sample, w_G is based on the Fama & French prior with $\sigma_\alpha = 4\%$ and the sample and w_r is the robust portfolio based on all CAPM and Fama & French model priors and the sample. We use j to refer to the acronyms u, c, d, F, G and r corresponding to the different models and $j = 0$ refers to the null model.

We are interested in comparing the average performance of the alternative portfolios and therefore we consider K bootstrap samples and compute the average ex-post performance of the alternative portfolios. Apart from the ex-post performance of each portfolio we store, among others the ex-ante expected performance according to each model prior, some portfolio characteristics and the ex-post expected loss. We use w_j^k, $\hat{\mu}_j^k$ and $\hat{\Sigma}_j^k$ to denote the optimal portfolio, estimated expected return vector and the estimated covariance matrix at bootstrap $k \in \{1, .., K\}$ under return model $j \in \{u, c, d, F, G\}$ and w_0 to denote the optimal portfolio under the null model. We report the statistics

[3] Available from the homepage of Professor K. French.

[4] Alternative values of γ affect the size of investment but do not affect the portfolio composition or, for robust portfolio choice, the decision whether or not to invest. Therefore the results for alternative values of γ will be scaled versions of the presented results.

ex-ante			
excess return	r_j^k	$=$	$w_j^{k\prime}\hat{\mu}_j^k$
variance	$(s_j^k)^2$	$=$	$w_j^{k\prime}\hat{\Sigma}_j^k w_j^k$
mean-variance utility	Q_j^k	$=$	$r_j^k - \frac{1}{2}\gamma {s_j^k}^2$
Sharpe ratio	Sh_j^k	$=$	r_j^k/s_j^k
ex-post			
excess return	$r_0(w_j^k)$	$=$	$\mu_0' w_j^k$
variance	$s_0^2(w_j^k)$	$=$	$w_j^{k\prime}\Sigma_0 w_j^k$
mean-variance utility	$Q_0(w_j^k)$	$=$	$r_0(w_j^k) - \frac{1}{2}\gamma s_0^2(w_j^k)$
Sharpe ratio	$Sh_0(w_j^k)$	$=$	$r_0(w_j^k)/s_0(w_j^k)$
expected loss	$L_0(w_j^k)$	$=$	$Q_0(w_0) - Q_0(w_j^k)$
portfolio statistics			
sum		$=$	$\iota' w_j^k$
norm		$=$	$\Vert w_j^k \Vert.$
cross-sectional st.dev.		$=$	$\sqrt{\Vert w_j^k\Vert^2/N - (\iota' w_j^k/N)^2}.$

For the robust portfolio w_R^k we (re)define the statistics

worst model(s)	j_{wc}^k	$=$	$arg\min_{j\in\{c,d,F,G\}} Q_j^k(w_R^k)$
expected mv	Q_R^k	$=$	$\min_{j\in\{c,d,F,G\}} Q_j^k(w_R^k)$
expected return	r_R^k	$=$	$\min_{j\in\{c,d,F,G\}} r_j^k(w_R^k)$
expected st.dev.	s_R^k	$=$	$\min_{j\in\{c,d,F,G\}} s_j^k(w_R^k)$
expected sharpe	Sh_R^k	$=$	$r_R^k/s_R^k.$

Table 3.1 reports the expected loss of the alternative portfolios relative to the optimal portfolio for the null model when the models are estimated on the entire Fama & French dataset. It provides a measure for model misspecification in the absence of estimation uncertainty. Naturally the unrestricted model has zero loss as it exclusively suffers from estimation uncertainty. The performance statistics corresponding to the weak model priors ($\sigma_\alpha = 4\%$) are similar to those of the unstructured model. On the other hand, the non-zero loss of the portfolios associated with strong model priors ($\sigma_\alpha = 0.1\%$) suggest that these models are misspecified. The Fama & French factor models suffer less model misspecification than the CAPM model. The Fama & French investor omits leverage but will use short positions to increase expected performance. The 'CAPM' typically predicts positive correlations for asset and market returns and refrains from leveraged portfolios as can be seen from the sum of portfolio weights which is lower than 1. The strong prior on the CAPM model, which is most restrictive, is also worst case in the robust approach. The robust portfolio choice and the portfolio corresponding to the strong CAPM prior are therefore similar.

Table 3.2 reports the averages and standard deviations of the above statistics for a bootstrap experiment with $K = 10000$ samples and $T = 60$ observations, hence with *estimation uncertainty*. The table also reports the fraction of bootstrap samples for which the corresponding portfolio is active, i.e. has a positive portfolio norm ('active'),

Table 3.1: Model misspecification

	unstructured	Fama & French		CAPM		robust
		$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 4\%$	$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 4\%$	
	w_u, w_0	w_F	w_G	w_c	w_d	w_r
expected performance (ex-ante)						
excess return (r)	4.62	1.08	4.55	0.16	4.37	0.16
standard deviation (s)	9.61	4.64	9.53	1.81	9.35	1.81
mean-variance utility (Q)	2.31	0.54	2.27	0.08	2.18	0.08
Sharpe ratio (Sh)	0.48	0.23	0.48	0.09	0.47	0.09
performance under null (ex-post)						
excess return (r)	4.62	1.10	4.56	0.17	4.43	0.17
standard deviation (s)	9.61	4.61	9.49	1.81	9.23	1.81
mean-variance utility (Q)	2.31	0.57	2.31	0.09	2.30	0.09
Sharpe ratio (Sh)	0.48	0.24	0.48	0.10	0.48	0.10
expected loss (L)	0.00	1.74	0.00	2.21	0.01	2.21
portfolio						
sum	1.64	0.72	1.62	0.39	1.59	0.39
cross-sectional stdev	1.23	0.16	1.21	0.03	1.15	0.03
stringent		0	0	1	0	

Notes: The table shows the ex-ante and ex-post performance of optimal mean variance portfolios ($\gamma = 5$) which are based on alternative return models corresponding to the acronyms u, c, d, F and G and the model robust portfolio w_R estimated on the entire Fama & French 25 portfolios dataset over the period July 1963 to December 2002. As estimated on the entire dataset, the results for the unstructured return model also present the results of the null model.

and the fraction of the bootstrap samples for which each return model is worst case for the robust portfolio ('stringent'). Moreover, $P(Q_0(w_j) \geq Q_j)$ and $P(Sh_0(w_j) \geq Sh_j)$ denote the fraction of the bootstraps for which the ex-post mean-variance performance and Sharpe ratio of portfolio w_j are at least their ex-ante expectations. The former statistic serves as measure of robustness of the expected performance.

The portfolios associated with little informative priors have worst ex-post performance. Apparently, the absence of model structuring leads to large estimation uncertainty in the return model parameters (compare to table 3.1) and leads to over-optimistic ex-ante expectations and suboptimal portfolios with a portfolio sum above one, hence leveraged. Jobson and Korkie (1981) and Michaud (1998) describe such portfolio allocations as error-maximizing portfolios.

The portfolios associated with strong prior beliefs in alternative structured return models, in this case the Fama & French model and CAPM model, have moderate risky asset allocations and have better ex-post performance. The structured return models decrease estimation uncertainty and, in terms of ex-post portfolio performance, this decrease in estimation uncertainty may outweigh the cost of a slight model misspecification.

The *model* robust portfolio which maximizes the worse case performance of alternative models c, d, F and G, has similar ex-post performance as the structured portfolios. Indeed, the robust portfolio mimics the most conservative and therefore typically the

Table 3.2: Portfolio performance of a model robust investor

	unstructured	Fama & French		CAPM		robust
		$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 4\%$	$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 4\%$	
	w_u	w_F	w_G	w_c	w_d	w_r
expected performance (ex-ante)						
excess return (r)	24.74	2.14	18.24	0.54	11.28	0.54
	(9.38)	(1.53)	(6.04)	(0.73)	(3.62)	(0.73)
standard deviation (s)	21.88	6.12	18.85	2.63	14.83	2.63
	(4.04)	(2.28)	(3.09)	(1.95)	(2.36)	(1.95)
mean-variance utility (Q)	12.37	1.07	9.12	0.27	5.64	0.27
	(4.69)	(0.77)	(3.02)	(0.37)	(1.81)	(0.37)
Sharpe ratio (Sh)	1.09	0.31	0.94	0.13	0.74	0.13
	(0.20)	(0.11)	(0.15)	(0.10)	(0.12)	(0.10)
performance under null (ex-post)						
excess return (r)	9.52	1.18	7.36	0.20	5.08	0.21
	(4.43)	(0.71)	(2.94)	(0.28)	(1.85)	(0.28)
standard deviation (s)	43.17	6.57	29.99	2.70	18.82	2.70
	(14.24)	(2.91)	(7.99)	(2.16)	(4.99)	(2.15)
mean-variance utility (Q)	-42.13	-0.12	-16.72	-0.10	-4.40	-0.09
	(35.45)	(0.72)	(11.75)	(0.31)	(4.10)	(0.31)
Sharpe ratio (Sh)	0.22	0.17	0.24	0.05	0.27	0.05
	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.08)
expected loss (L)	44.44	2.42	19.03	2.41	6.70	2.40
	(35.45)	(0.72)	(11.75)	(0.31)	(4.10)	(0.31)
ex-ante to ex-post						
$P(Sh_0(w_j) \geq Sh_j)$	0.0	11.3	0.0	27.7	0.0	28.2
$P(Q_0(w_j) \geq Q_j)$	0.0	10.9	0.0	27.4	0.0	28.0
portfolio						
active	1.00	1.00	1.00	1.00	1.00	1.00
sum	3.26	0.81	2.59	0.44	1.94	0.44
cross-sectional stdev	6.16	0.21	3.99	0.04	2.20	0.04
stringent		0.68	0.43	0.997	0.01	

Notes: The table shows the average ex-ante and ex-post performance of optimal mean variance portfolios ($\gamma = 5$) which are based on alternative return models corresponding to the acronyms u, c, d, F and G and the model robust portfolio w_R over 10.000 bootstrap samples of 60 random observations from the Fama & French 25 portfolios dataset over the period July 1963 to December 2002. The ex-post performance is evaluated on the null model. All numbers are averages (stdev's) over the bootstraps, and are, except the Sharpe ratios and portfolio characteristics, given in percentages per month.

most structured portfolio which is, in more than 99% of the bootstrap samples, the portfolio based on the strong CAPM prior. The crucial dependence of portfolio choice under uncertainty on the most restrictive return model prior is also observed for a Bayesian approach (Lindleys paradox).

The reliability, measured by the fraction of bootstrap samples for which the portfolio attains, ex-post, the ex-ante expected performance, is too small to label the ex-ante performance estimates as robust. Estimation uncertainty and, for the robust approach, a limited number of alternative models lead to ex-ante overestimation of the performance.

We are also interested in the tolerance of the portfolios when one of the alternative prior return models is correct. We therefore repeat the experiment under alternative instances of the true return model μ_0, Σ_0, in particular the models that correspond to a strong belief in the CAPM or the Fama & French model. We change the original null model gradually, as measured by a scalar $\delta \in [0, 1]$, to any of the structured asset pricing models. When $\delta = 0$ we maintain the original dataset; when $\delta = 1$ the dataset corresponds with a sample from the alternative model. To preserve the random properties of the dataset other than implied by the model, we consider the standardized residuals ε_t, which follow from

$$y_t = \mu_0 + C\varepsilon_t$$

where C is the lower triangular part of the Choleski decomposition of Σ_0, as characteristic for the dataset. We build the new dataset on these residuals and the intended alternative factor model j estimated on the entire original Fama & French dataset with parameter estimates α_j, B_j and D_j. The alternative model leads to an alternative dataset with 'observations'

$$\tilde{y}_t = \alpha_j + B_j x_t + C_j \varepsilon_t$$

where C_j is the lower triangular part of the Choleski decomposition of D_j. The new dataset, indicated by a change coefficient δ, is

$$y_t \rightarrow (1 - \delta)y_t + \delta\tilde{y}_t.$$

Figure 3.6 shows the expected loss of the portfolios when we change the dataset towards the Fama & French or CAPM model. Naturally the portfolio performance of the intended alternative model improves when δ increases. More interesting is the performance of the other portfolios. The robust portfolio crucially depends on the most restrictive model, which results from a strong CAPM prior. The performance of the CAPM model improves when we change to a dataset which features the structure of the Fama & French or CAPM model. Hence the robust approach also performs well under such changes. On the other hand, the unrestricted portfolio hardly exploits the change in return structure and remains suboptimal.

We also study the change in portfolio performance when uncertainty decreases due to an increase in the number of observations. Figure 3.5 shows the portfolio performance for different numbers of observations. The performance of the portfolios associated with the unstructured return model naturally improves as estimation uncertainty decreases and will eventually converge to zero expected loss. The performances of the portfolios associated with weak return model priors shows a similar behavior. On the other hand, the performance of the portfolios associated with strong return model priors improves much slower. The effect of the prior is strong relative to the information in the data. Therefore the model misspecification in the prior is perceptible even when the number of observations is large. The robust portfolio which crucially depends on the CAPM model, reacts similarly to an increase in the number of observations.

3.4 Model and Estimation Uncertainty

In this section we extend the robust approach by adding estimation uncertainty. In addition to reporting (μ_j, Σ_j), each expert now also provides the decision maker with her uncertainty set (in essence a confidence interval) around these estimates. The uncertainty set $\mathcal{U}_j$ of a particular expert j will contain the parameter values that are, in the expert's believe, plausible. We confine to experts who exclusively consider uncertainty in the estimator for the expected return[5].

Investors who are confident that a particular expert j is correct and want to be estimation robust use $\mathcal{U} = \mathcal{U}_j$ and solve

$$\max_{w} \min_{\mu \in \mathcal{U}} Q_j(\mu, w) \tag{48}$$

with

$$Q_j(\mu, w) = \mu' w - \frac{1}{2}\gamma w' \bar{\Sigma}_j w$$

The specific form of the uncertainty set depends on the expert. Common in their approaches is that they derive this set from the posterior distribution of the parameter for each of the models. The next subsection elaborates on the construction of the uncertainty set for alternative experts.

Apart from investors who base their portfolio on one of the expert models $j \in \{1, .., J\}$, we also consider an investor who adopts a robust approach to the alternative expert models. The model- *and* estimation robust investor solves

$$\max_{w} \min_{j} \min_{\mu \in \mathcal{U}_j} Q_j(\mu, w). \tag{49}$$

Note that the investor now faces a larger set of possible return models. The robust decision will thus be more conservative than before.

3.4.1 Estimation uncertainty

Consider an expert who adopts a non-informative prior on the expected returns and suppose Σ is known. The uncertainty of the expected excess return vector μ conditional on the observation $\bar{y}$, may be described by the uncertainty matrix $\Omega = \Sigma/T$ and

$$(\mu - \bar{y})'\Omega^{-1}(\mu - \bar{y}) \sim \chi^2(n).$$

[5] Errors in means are more critical than errors in variances, and errors in variances are more critical than errors in covariances (see Chopra and Ziemba (1993))

Consequently the uncertainty set

$$\mathcal{U} = \{\mu : (\mu - \bar{y})'\Omega^{-1}(\mu - \bar{y}) \leq \theta^2\},$$

will contain μ with sufficient probability (95%) if $\theta^2 = \chi^2_{inv}(N, 0.95)$.

Alternatively, consider a factor model (36)-(37). From (41) and (42) it follows

$$\begin{aligned}\Omega_a &= Var(a|X,\ Y) = \tilde{D} \otimes F^{-1} \\ \Omega_\nu &= Var(\nu|X,\ Y) = \hat{\Psi}/(T - K - 2)\end{aligned}$$

and B and ν are independent in the posterior. Following a similar argumentation as for (3.4.1) we define the uncertainty sets for a and ν,

$$\begin{aligned}\mathcal{U}_a &= \{(a - \tilde{a})'\Omega_a^{-1}(a - \tilde{a}) \leq \theta_\beta^2\} \\ \mathcal{U}_\nu &= \{(\nu - \tilde{\nu})'\Omega_\nu^{-1}(\nu - \tilde{\nu}) \leq \theta_\nu^2\}.\end{aligned}$$

We choose θ_β and θ_ν such that the uncertainty sets cover 95% posterior probability[6].

Given a portfolio w, the innermost minimization in (49) reduces to

$$\begin{aligned}\min_{\alpha, B,\ \nu} w'(\alpha + B\nu) - \frac{1}{2}\gamma w'(B\tilde{\Psi}B' + \tilde{D})w \\ vec([\alpha, B]') \in \mathcal{U}_a \\ \nu \in \mathcal{U}_\nu\end{aligned} \tag{50}$$

Unfortunately the uncertain parameters enter both in a bi-linear and quadratic term of the objective (50) and we are not able to solve this bi-linear optimization problem. Only if we abstract from uncertainty in one of the two estimators and ignore the effect of B on the variance term, we can solve the problem exactly. As an alternative we consider the uncertainty in the expected return directly and apply the first-order Taylor expansion for standard errors on μ,

$$\begin{aligned}\tilde{\mu} - \mu &\approx (\tilde{\alpha} - \alpha) + (\tilde{B} - B)\tilde{\nu} + \tilde{B}(\nu - \tilde{\nu}) \\ &= \tilde{B}(\nu - \tilde{\nu}) + I_N \otimes \begin{pmatrix} 1 \\ \tilde{\nu} \end{pmatrix}' (a - \tilde{a})\end{aligned}$$

and consequently

$$\Omega = \mathsf{E}[(\mu - \tilde{\mu})(\mu - \tilde{\mu})'] = \frac{1}{T}\tilde{B}\tilde{\Psi}\tilde{B}' + \begin{pmatrix} 1 \\ \tilde{\nu} \end{pmatrix}' F^{-1} \begin{pmatrix} 1 \\ \tilde{\nu} \end{pmatrix} \tilde{D}$$

[6] This Carthesian product of uncertainty sets is not exactly the highest posterior density region associated with 0.95^2 cumulative posterior density but convenient for computations.

We derive an uncertainty set of the form

$$\mathcal{U} = \{\mu : (\mu - \tilde{B}\tilde{\nu})'\Omega^{-1}(\mu - \tilde{B}\tilde{\nu}) \leq \theta^2\}, \tag{51}$$

and we use[7] $\theta^2 = F_{inv}(K, T, 0.95)$ if $\sigma_\alpha \approx 0$ and $\theta^2 = \chi^2_{inv}(N, T, 0.95)$ otherwise to have a reasonable estimate of θ which leads to an uncertainty set with 95% posterior density, though not necessarily the highest posterior density region.

3.5 Model and Estimation robust: Empirical study

A study on the effects of estimation robust portfolio choice allows us to compare the importance of a robust approach to model uncertainty and estimation uncertainty. Table 3.1 already reported the expected loss due to model misspecification for each structured model and table 3.2 shows the additional effect of estimation uncertainty for each model. The results of this section will show to what extent the effect of estimation uncertainty shown in table 3.2 can be resolved by an estimation (uncertainty) robust approach.

We repeat the bootstrap of section 3.3 but now consider the optimal portfolios w_j^k for $j = \{u, c, d, F, G\}$ which solve (48), with $\hat{\mu} = \hat{\mu}_j^k$, $\hat{\Sigma} = \tilde{\Sigma}_j^k$ and $\Omega = \Omega_j^k$. We choose $\Omega = \tilde{\Sigma}/T$ and as Ω is an estimate we let $\theta^2 = T^2_{inv}(0, 95, N, T-1) = \frac{N(T-1)}{T-N}F_{inv}(0.95, N, T-N)$. The robust expected return which is corrected for estimation uncertainty replaces the naive expected return,

$$r_j^k(w_j^k) = w_j^{k\prime}\hat{\mu}_j^k - \theta\sqrt{w_j^{k\prime}\hat{\Omega}w_j^k}.$$

The robust portfolio w_R^k solves (49). Again we evaluate all portfolios on their (ex-post) performance under the null model.

Table 3.3 reports the performance of estimation robust portfolios for a bootstrap experiment with $K = 10.000$ samples with 60 observations each. The robust approach to uncertainty reduces the expected loss of the portfolio associated with the unstructured return model considerably. More specifically, a robust approach to estimation uncertainty leads to an ex-post performance which is comparable to the performance achieved by prominent (structured) return models. Both approaches decrease the exposure to estimation uncertainty: the structured return models by a priori structuring the return model such that the effects of uncertainty are limited, the robust approach by evaluating the uncertainty.

The factor models also profit from an estimation robust approach, albeit to less extent. All portfolios become more conservative compared to a naive approach to estimation uncertainty. The model robust approach which considers all structured models and the parameter uncertainty associated with each model almost stays out of the risky market.

[7]Ideally θ corresponds to the 95th percentile of $(\mu - \tilde{B}\tilde{\nu})'\Omega^{-1}(\mu - \tilde{B}\tilde{\nu})$ which we are unable to compute directly. The reported values are reasonable approximates to the desired percentile.

Table 3.3: Portfolio performance of a model- and estimation robust investor

	unstructured	Fama & French		CAPM		robust
		$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 2\%$	$\sigma_\alpha = 0.1\%$	$\sigma_\alpha = 2\%$	
	w_u	w_F	w_G	w_c	w_d	w_r
expected performance (ex-ante)						
excess return (r)	0.24	0.03	0.97	0.01	0.13	0.00
	(0.87)	(0.16)	(1.36)	(0.07)	(0.33)	(0.05)
standard deviation (s)	0.89	0.24	3.39	0.08	0.81	0.05
	(1.99)	(0.73)	(2.81)	(0.42)	(1.36)	(0.29)
mean-variance utility (Q)	0.12	0.01	0.48	0.00	0.06	0.00
	(0.44)	(0.08)	(0.68)	(0.03)	(0.16)	(0.02)
Sharpe ratio (Sh)	0.04	0.01	0.17	0.00	0.04	0.00
	(0.10)	(0.04)	(0.14)	(0.02)	(0.07)	(0.01)
performance under null (ex-post)						
excess return (r)	0.47	0.05	1.42	0.01	0.32	0.01
	(1.20)	(0.18)	(1.42)	(0.05)	(0.58)	(0.05)
standard deviation (s)	2.05	0.28	5.65	0.09	1.12	0.05
	(4.97)	(0.89)	(5.13)	(0.48)	(1.99)	(0.34)
mean-variance utility (Q)	-0.25	0.03	-0.03	0.00	0.18	0.01
	(2.52)	(0.10)	(1.46)	(0.02)	(0.30)	(0.03)
Sharpe ratio (Sh)	0.22	0.17	0.24	0.11	0.27	0.24
	(0.07)	(0.06)	(0.07)	(0.10)	(0.06)	(0.07)
expected loss (L)	2.56	2.27	2.34	2.30	2.12	2.30
	(2.52)	(0.10)	(1.46)	(0.02)	(0.30)	(0.03)
ex-ante to ex-post						
$P(Sh_0(w_j) \geq Sh_j)$	92.5	96.6	71.5	84.3	98.3	99.5
$P(Q_0(w_j) \geq Q_j)$	89.4	87.7	66.4	76.1	97.7	99.7
portfolio						
active	0.29	0.17	0.86	0.08	0.43	0.04
sum	0.16	0.04	0.51	0.02	0.13	0.01
cross-sectional stdev	0.29	0.01	0.74	0.00	0.13	0.00
stringent		0.04	0.02	0.03	0.00	

Notes: The table shows the average ex-ante and ex-post performance of optimal, estimation robust, mean variance portfolios ($\gamma = 5$) which are based on alternative return models corresponding to the acronyms u, c, d, F and G and the model robust portfolio w_R over 10.000 bootstrap samples of 60 random observations from the Fama & French 25 portfolios dataset over the period July 1963 to December 2002. The ex-post performance is evaluated on the null model. All numbers are averages (stdev's) over the bootstraps, and are, except the Sharpe ratios and portfolio characteristics, given in percentages per month.

An estimation robust approach considerably improves the reliability, measured by the fraction of bootstrap samples for which the portfolio attains, ex-post, the ex-ante expected performance. The reliability of a naive approach to estimation uncertainty is, at best, 30%. The reliability of estimation robust portfolios is typically around 90%. This illustrates the crucial importance to take estimation uncertainty into consideration to form reliable ex-ante performance estimates. The observed reliability also shows that the characterization of estimation uncertainty (uncertainty sets) is adequate to provide sufficient robustness[8].

The gain in robustness compared to portfolios which are not estimation robust, is partly achieved by passive strategies that result when the T observations of the bootstrap sample do not support investment opportunities with significant positive performance. The density plot, figure 3.4, of the portfolio performances support this explanation. The spike at a loss of 2.31% corresponds to the fraction of passive portfolios. The active portfolios have a loss that is typically below 2.31%. The truncation of the distribution suggests that a robust approach to estimation uncertainty succeeds in filtering the samples that produce unreliable parameter estimates and would lead to portfolios with worse performance than a passive portfolio.

A robust approach to estimation uncertainty leads to smaller risky investment. The unstructured, estimation robust portfolio has, averaged over the *active* investments, a portfolio sum of 0.53 (= 0.16/0.30) and a cross-sectional standard deviation of 0.97 (= 0.29/0.30).

Although model structuring typically leads to smaller estimation uncertainty, we observe that the portfolio based on a strong prior believe in the Fama & French model is less active than the portfolio based on the unstructured model. Apparently, the Fama & French prior leads to sober predictions for the parameters such that the uncertainty set for the parameter, though small, contains parameters that imply a larger loss than a passive strategy. We also observe that, when estimation robust, not the CAPM model but the Fama & French model is worst case to the robust portfolio. This implies that the Fama & French model, which is as regards model misspecification less restrictive than the CAPM, suffers larger estimation uncertainty than the CAPM model. As a result the worst case according to the Fama & French model will be most restrictive.

The improvement in portfolio performance due to a robust approach to estimation uncertainty is larger than the improvement due to a model robust approach. This observation differs from the conclusion of Avramov (2002) who studies a Bayesian decision maker. He concludes that accounting for model uncertainty is more important than considering estimation uncertainty. An important difference between the study by Avramov (2002) and this study is the diversity of (structured) return models which is considered. Naturally the effect of a robust approach increases when we include models which lead to a greater dispersion of expected return forecasts. This could lead to an improvement in the performance of a model robust approach. It would be interesting to study the robust approach in the context considered by Avramov (2002). One could use the posterior model probabilities to select the alternative models for the uncertainty set.

Figure 3.5 shows the portfolio performance for various degrees of uncertainty. The performance of a structured portfolio improves only slowly when the number of observations increases and consequently uncertainty decreases. As the number of observations

[8]The non-normality of the data, uncertainty in the covariance matrix, possible model misspecification and first order approximations to derive the uncertainty set may cause the slight deviations of the targeted 95% robustness.

increases, the estimate of μ converges slowly towards the unrestricted mean. But the effect of the prior dominates the information in the data and therefore the consequences of model misspecification (see table 3.1) are present even for a large number of observations.

On the other hand, the portfolio based on an uninformative return model prior becomes active more often and its performance improves considerably. The uncertainty set shrinks as the number of observations increases and eventually converges to the true parameter value. Hence for a sufficient number of observations the ex-post expected loss of this portfolio will be zero.

We observe that estimation robust portfolio choice based on an uninformative prior has some desirable properties. When the number of observations is small and estimation uncertainty is large, it attains the same performance as portfolios which use return model structure to decrease estimation uncertainty. When uncertainty decreases the portfolio converges, faster than structured portfolios, to the truly optimal portfolio.

3.6 Conclusion

We have shown that the model robust portfolio is the optimal mean-variance portfolio corresponding to a, endogenously determined, combination of alternative return models. For some special cases we also showed that the robust portfolio is a convex combination of the optimal portfolios corresponding to the alternative models. In that case, the robust portfolio has smaller expected loss than the worst of the alternative portfolios. Moreover, the robust portfolio has least expected loss for a range of models in-between the alternative models. When we consider that experts propose sensible return models, then alternative expert's beliefs typically circumscribe a set of models which contains the true model. In this event, the robust portfolio is likely to outperform the optimal portfolios associated with alternative models.

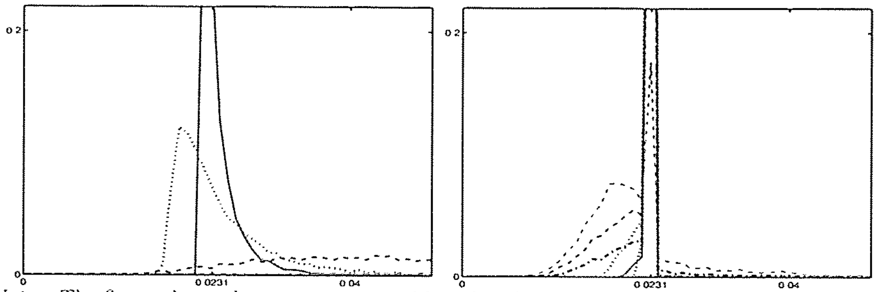
We considered robust portfolio choice in an empirical setting with alternative models that are based on strong and weak beliefs in alternative asset pricing models. We observe that the model robust portfolio resembles closely the optimal portfolio associated with the model that is worst to the robust portfolio. This worst case model is typically the most restrictive model: the CAPM if estimation uncertainty is not considered, the Fama & French model for a robust approach to estimation uncertainty.

Empirical results show favorable results for the estimation robust portfolio based on a non-informative return model prior: When the number of observations is small and estimation uncertainty is large, it attains the same performance as portfolios that use return model structure to decrease estimation uncertainty. When uncertainty decreases, the portfolio converges, faster than the structured portfolio, to the truly optimal portfolio.

In this and the previous chapter, we assumed that expected returns are constant over time. To get around this implausible assumption, we rely on the 60 most recent monthly observations to estimate current expected returns. In chapter 4 we relax the assumption that expected returns are constant over time. Time variation in expected returns offers opportunities to investors. It also leads to more uncertainty about the specification of the return generating process: Predictability brings higher utility to the investors but uncertainty about the predictability may reduce the gains from the perspective of robustness.

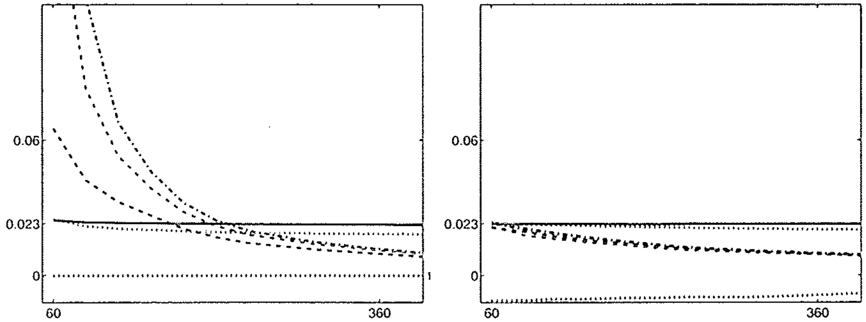
A.3 Figures

Figure 3.4: Expected loss distribution.



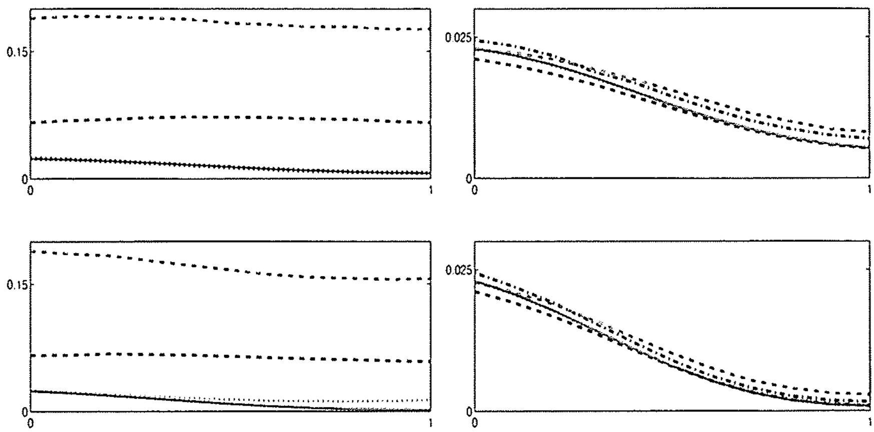
Notes: The figure shows the ex-post expected loss distribution over 10.000 bootstrap samples of 60 random observations from the Fama & French 25 portfolios dataset over the period July 1963 to December 2002. The left panel concerns portfolios which ignore parameter uncertainty. The right panel considers estimation robust portfolios. The different lines correspond to alternative model priors: unstructured model (dash-dot), weak Fama & French prior (dashed grey), weak CAPM prior (dashed), strong Fama & French prior (dotted grey), strong CAPM prior (dotted) and the model robust portfolio (solid line).

Figure 3.5: Expected loss as a function of uncertainty



Notes: The figure shows the portfolio performance in terms of expected loss (left axis) as a function of the number of observations, when the investor ignores estimation uncertainty (left-panel) and when the investor is robust to estimation uncertainty (right-panel). The right axis and the dotted line on the bottom measure the fraction of bootstraps for which the model (and estimation) robust investor is active. The different lines correspond to alternative model priors: unstructured model (dash-dot), weak Fama & French prior (dashed grey), weak CAPM prior (dashed), strong Fama & French prior (dotted grey), strong CAPM prior (dotted) and the model robust portfolio (nearly horizontal solid line).

Figure 3.6: Expected loss for alternative null models



Notes: The figure shows the portfolio's average ex-post expected loss (left axis) when the null model gradually changes to a CAPM model (upper panel) and a Fama & French model (lower panel), for portfolios which are not robust to estimation uncertainty (left-panels) and for estimation robust portfolios (right-panels). The lines correspond to portfolios based on alternative model priors: unstructured model (dash-dot), weak Fama & French prior (dashed grey), weak CAPM prior (dashed), strong Fama & French prior (dotted grey), strong CAPM prior (dotted) and the model robust portfolio (solid line). The portfolio associated with the unstructured return model is not depicted in the left panels as it is a horizontal line outside the depicted range.

Chapter 4

Uncertainty in dynamic portfolio problems

Life can only be understood backwards;
but it must be lived forwards.
Søren Kierkegaard (1813-1855).

In this chapter we study the decisions of a robust investor with regard to multi-period portfolio choice in the presence of return predictability and parameter uncertainty. We use the same context as Barberis (2000), who considers optimal multi-period portfolio choice with a single risky asset and a return model which features time-varying expected returns. We compare the robust investor to a naive investor who ignores parameter uncertainty and an investor who adopts a Bayesian approach to decision making under uncertainty. Moreover we study the relations between return predictability, parameter uncertainty and robustness and consider whether active participation in the risky asset market which emanates from return predictability (in the estimated model) is sustainable if the investor aims to be robust. We also characterize the worst plausible return model for the robust solution.

Multi-period portfolio choice differs from (a sequence of) one-period portfolio choice problems as the investor may anticipate the possibility of future corrective actions based on concurrent information in her current decision. The dynamics of the stock return process (e.g. mean reversion) and return predictability may favor a dynamic investment strategy which adapts to changes in the investment opportunity set over time. On the other hand, uncertainty about the stock return model questions the estimated return dynamics and presence of return predictability and thereby restricts the extent to which we should adapt the investment strategy to an estimated stock process. A robust investor deals with uncertainty by using an investment strategy that has best worst case performance over some set of plausible return models.

A solution to a multi-period portfolio choice problem is a contingency plan of portfolio allocations at every instant over the time horizon and for each economic state. Solving multi-period portfolio choice problems is difficult due to uncertainty about the future economy, the possibility to adapt future decisions to new information and inter-temporal relations between decisions. The success of providing an empirical relevant study on portfolio optimization, hinges on our ability to preserve the key properties of the original problem, while simplifying the model to an extent that allows us to solve the model.

Barberis (2000) proposes a solvable and relevant multi-period portfolio choice problem with a finite number of re-balancing points, a finite state space that describes the (uncertain) development of the economy and a problem structure that enables us to use the Bellman principle of optimality for dynamic decision problems. The restriction on re-balancing can be justified as it corresponds to practice. A finite state space can only approximate the original problem, but if accurate enough, it still provides the correct intuition for the solutions.

4.1 Predictability, uncertainty and robustness

This section presents the context of the portfolio choice problem that we use to study the behavior of naive, Bayesian and robust investors in a dynamic environment. We define the investors objective and the investment opportunity set available to the investor for maximizing the objective.

We consider a multi-period portfolio choice problem with a finite horizon set to T months. The investor's preferences are given by a utility function $u(\cdot)$. Utility due to intermediate consumption is disregarded and utility depends exclusively on the terminal wealth W_T. We consider investors with a utility function that features constant relative risk aversion power utility over terminal wealth:

$$u(W_T) = \frac{W_T^{1-\gamma}}{1-\gamma}. \tag{1}$$

Terminal wealth accumulates through portfolio returns over time. If wealth at time t is W_t, and w_t is the asset allocation over the period $[t, t+1)$, wealth at time $t+1$ is given by

$$W_{t+1} = (1 + R_{t+1}(w_t))W_t \tag{2}$$

where w_t denotes the investment in the risky asset with an implicit investment of $(1-w_t)$ in the riskfree asset and $R_{t+1}(w_t)$ denotes the portfolio return on asset allocation w_t over period $[t, t+1)$. We consider portfolio choice which is limited to two assets: Treasury bills and a stock index. We stylize the problem by assuming that the continuously compounded monthly return on Treasury bills is a constant r_f. The returns for the stock index are given by a model for the monthly continuously compounded excess stock index returns y_t:

$$y_t = \alpha_1 + \beta_1 x_{t-1} + \varepsilon_{1t}. \tag{3}$$

with the dividend yield x_t serving as a predictor variable. To complete the model for long-run returns, we specify a process for the predictor variable:

$$x_t = \alpha_2 + \beta_2 x_{t-1} + \varepsilon_{2t}. \tag{4}$$

The error terms $(\varepsilon_{1t}, \varepsilon_{2t})$ have a joint normal distribution with covariance matrix Σ. Note that $\beta_1 = 0$ would reduce the model to a static return model.

A risky (stochastic) stock index return makes wealth accumulation (2) a stochastic process. Indeed the portfolio return over the period $[t, t+1)$ on asset allocation w_t is

$$1 + R_{t+1}(w_t) = (1 - w_t)e^{r_f} + w_t e^{r_f + y_{t+1}}. \tag{5}$$

The investor maximizes expected terminal utility (1) conditional on x_0 (indicated by subscript 0):

$$\max_{w \in \mathcal{W}} \mathsf{E}_0[u(W_T)], \tag{6}$$

subject to the wealth accumulation constraint (2). In our study the set of feasible investment strategies is $\mathcal{W} = \{w : \ 0 \leq w \leq 1\}$.

4.1.1 Uncertainty

The return model (3)-(4) is estimated to predict future excess returns. When the return model is correctly specified, all deviations from the correct return model are exclusively due to estimation errors in the estimators for the parameters $\xi = [\alpha_1 \ \beta_1 \ \alpha_2 \ \beta_2]'$ and Σ. To simplify the analysis we focus on estimation errors (uncertainty) in the parameters[1] ξ. Possible estimation errors induce uncertainty about all quantities that depend on the parameters.

The expectation operator and the portfolio return (5) are functions of the parameter ξ. If we use an estimate for ξ which may, due to estimation uncertainty, deviate from the true parameter value, we introduce uncertainty in the expectation operator and the portfolio return.

When required for clarity, we will explicitly show the functional dependence of the expectation operator and portfolio return on the parameter by enclosing ξ in the notation.

Conditional on an estimate for the model parameters $\hat{\xi}$, a *naive* investor uses the return model (3)-(4) to form a conditional distribution for future stock returns and solves the problem

$$\begin{array}{ll} \max_{w \in \mathcal{W}} & \mathsf{E}_0[u(W_T) | \ \hat{\xi}] \\ \text{subject to} & W_{t+1} = (1 + R_{t+1}(w_t, \hat{\xi}))W_t. \end{array} \tag{7}$$

[1] Σ can be estimated much more precisely than ξ.

A *Bayesian* investor recognizes that an estimated model is only approximately true. Therefore she does not condition on $\hat{\xi}$, but integrates over all possible parameter values using the posterior parameter distribution. Given $\tilde{T}$ observations on monthly dividend yield $x_{-1}, .., x_{-\tilde{T}}$ ($X = [\iota\ x]$), monthly continuously compounded excess returns $Y = [y_{-1}, .., y_{-\tilde{T}}]'$ and the posterior parameter distribution $p(\xi |\ Y,\ X)$, the Bayesian investor maximizes

$$\mathsf{E}_{B,0}\left[u(W_T)\right] = \int \mathsf{E}_0\left[u(W_T) \mid \xi\right] p(\xi \mid X, Y) d\xi \tag{8}$$

with wealth accumulation following from $W_{t+1} = (1 + R_{t+1}(w_t, \xi))W_t$. We will follow Barberis (2000) and consider a Bayesian investor who uses a standard non-informative prior.

Also a *robust* investor is aware that an estimated model is only approximately true. The investor wishes to account for this uncertainty. Unlike a Bayesian investor, however, she does not (know how to) assign probabilities to alternative models. Instead she considers the least favorable alternative ξ from within some uncertainty set $\mathcal{U}$ of conceivable parameter values and maximizes,

$$\min_{\xi \in \mathcal{U}} \mathsf{E}_0\left[u(W_T) \mid \xi\right]. \tag{9}$$

The investor is confident that the true model parameters lie within the uncertainty set. Therefore the investor trusts that an asset allocation with good performance for the least favorable model in the set will also perform well for the true model. The critical determinant of robust decision making is the set of alternative models considered by the investor. If this set is large, the worst case return model and consequently the investor's behavior will be rather pessimistic and robustness will result in passive portfolio strategies. On the other hand, if the investor has a strong belief in some reference model and the set of alternatives is small, her portfolio will be almost equal to the portfolio of a naive investor who relies on the reference model.

We use the return model (3)-(4) estimated on the historical observations as reference model denoted with parameter (estimate) $\hat{\xi}$. We form the uncertainty set in a similar way as a confidence region for the true parameter values and consider the set of parameter values with highest posterior density with a cumulative density of, say, 95%, given the observed data,

$$\mathcal{U} : \int_{\xi \in \mathcal{U}} p(\xi | X,\ Y) d\xi = 0.95 \tag{10}$$

Assume Σ is known, and note that the conditional distribution of ξ is multivariate normal, hence a member of the class of ellipsoidal distributions. The equi-probable parameter values form an ellipsoid and the uncertainty can be expressed as

$$\mathcal{U} = \{\xi :\ (\xi - \hat{\xi})'\Omega^{-1}(\xi - \hat{\xi}) \leq \theta^2\}. \tag{11}$$

We calibrate the uncertainty set on the posterior distribution of Barberis (2000) which is based on an uninformative prior. In this case the covariance matrix $\Omega = \Sigma \otimes (X'X)^{-1}$ and $\theta = \sqrt{\chi^2_{inv}(4)}$.

Note that if the static model ($\beta_1 = 0$) is part of the uncertainty set $\mathcal{U}$, return predictability is not significant according to the robust investor. In this case the robust portfolio choice must consider predictable and non-predictable returns.

Given the uncertainty set, a robust investor chooses an asset allocation that is (a) *feasible* for all parameters, i.e. an investment strategy that satisfies the budget constraint for all parameters $\xi \in \mathcal{U}$, and (b) has best robust performance, which means that the worst performance over the parameter set is maximal. The *robust buy-and-hold portfolio choice problem* is

$$\begin{aligned} &\max_{w \in \mathcal{W}} \min_{\xi \in \mathcal{U}} \; \mathsf{E}_0[u(W_T(w)|\;\xi)] \\ &\text{subject to } W_T \leq (1 + R_T(w, \xi))W_0, \quad \forall \xi \in \mathcal{U}. \end{aligned} \tag{12}$$

A solution to (12) is an asset allocation w with the highest level of expected utility which is robust to the parameter values in the uncertainty set.

The robust model of the dynamic multi-period problem is more intricate. Nilim and El Ghaoui (2002) propose a model and solution method to the robust multi-period dynamic programming problem. If we apply their model to problem (6), we obtain the robust optimization problem

$$\begin{aligned} &\max_{w,\; W} \min_{\xi \in \mathcal{U}} \; \mathsf{E}_0[u(W_T(w)|\;\xi)] \\ &\text{subject to } W_{t+1} \leq (1 + R_{t+1}(w_t, \xi_t))W_t, \quad \forall \xi_t \in \mathcal{U},\; t \in \{1, .., T-1\}. \end{aligned} \tag{13}$$

A feasible solution to (13) consists of portfolio allocations $w = \{w_t\}_{t=1}^{T-1}$ and wealth variables $W = \{W_t\}_{t=2}^{T}$ that satisfy the constraints for each period and for all parameter values.

However, we are dealing with a special problem for which model (13) may lead to solutions that are more conservative than necessary. The worst case behavior in model (13) may result from choosing different worst case parameters for alternative time periods: Indeed if we consider two constraints in (13) corresponding to two arbitrary time periods t_1 and t_2, both constraints must hold for all parameters in the uncertainty set and in particular for the worst case parameter. It is quite possible that the worst case parameter configurations for the constraint at t_1 and t_2 differ. Yet, in our problem setting, the choice for a parameter value fixes the entire stochastic process, for all periods (or in finite state space, it fixes the scenario tree). A parameter that is fixed over time leads to inter-temporal dependencies. Let us illustrate this with an example.

Consider a two period return model that is stripped from all uncertainty except parameter uncertainty. The parameter may take values ξ_a or ξ_b. Under ξ_a, the return in the first and second period is 0% and 10% respectively. Under ξ_b, the return in the first and second period is 10% and 0% respectively. Both parameters imply a two-period return of 10%. Nevertheless, according to the constraints in model (13), we must reckon with the worst case return of 0% in the first period, due to ξ_a and similarly a worst case return of 0% in the second period, due to ξ_b. Obviously, this is more conservative than the two period robust return for investment in the risky asset.

The crucial difference between our problem and model (13) is the function of the variables W_t. In our problem definition, these merely serve as 'book-keepers' and are not subjected to other constraints such that only terminal wealth W_T is relevant for the problem. However in model (13), the W_t serve as decision variables that must be guaranteed, also at intermediate time periods. As seen by the example where robust return was zero instead of 10%, this leads to a 'leak' in the wealth accumulation.

Intermediate outcomes would be relevant if the utility function would contain consumption in every period as in Campbell and Viceira (2002). This would be a more difficult problem to solve.

We resolve this modelling issue by substituting for the redundant state variable wealth and express terminal wealth as a function of the asset allocations,

$$W_T(w, \xi) = \prod_{t=1}^{T} \left(1 + R_t(w_{t-1} |\, \xi)\right) W_0. \tag{14}$$

We substitute (14) in (6) and omit the variables W_t in our problem definition. Accordingly, we define the *robust multi-period portfolio choice problem* as:

$$\max_{w \in \mathcal{W}} \min_{\xi \in \mathcal{U}} \mathsf{E}_0 \left[u(W_T(w)) |\, \xi \right], \tag{15}$$

with W_T defined in (14). A robust solution is a plan which specifies the portfolio allocation for every time period and (future) economic state and this plan maximizes the worst case expected utility.

4.1.2 Portfolio dynamics

The extent to which the portfolio choice problem is dynamic depends, apart from the return dynamics, on the frequency of the asset allocation decisions. If the investor is confined to a *buy-and-hold strategy*, she chooses an allocation at the beginning of the horizon and passively awaits the results of her decision at the end of the horizon. Although this involves no dynamic decision making, a sensible investment strategy anticipates the future development of the dynamic return model. A *dynamic strategy* allows the investor to rebalance her portfolio at certain time points, say once a year. In this case, the investor chooses her current allocation, with the prospect that she may rebalance at the begining of every year to come, taking into consideration the information that will be available at that time. Indeed, a number of rebalancing possibilities allows her to react timely to new developments with so-called recourse actions. With the knowledge that she can alter her strategy to adapt to the return model dynamics and can possibly correct her decisions later if the economy turns against her, the investor may wish to choose an initial asset allocation that differs from a buy-and-hold strategy.

4.2 Investment opportunities

In this section we analyze the data that we will use as input for portfolio choice in the next sections. We consider the uncertainty in the data and its potential to sustain a positive Sharpe ratio which is an indicator for the profitability of the investment opportunity set.

4.2.1 Data

The estimation results by Barberis (2000), for convenience repeated in table 4.1, serve as data input. Barberis (2000) considers two samples of different length. The first sample, consists of 523 monthly observations of stock index returns, dividend yields and Treasury bill returns from June 1952 to December 1995. The second sample has 120 observations from January 1985 to December 1995. The stock index return and Treasury bill returns are used to calculate the excess stock index return. The continuously compounded monthly risk-free rate is set to 0.36%.

The estimation results of the two samples differ fundamentally in parameter uncertainty. The first sample produces a significant mean excess return, the second sample has a mean excess return which is not significantly different from zero. Consequently, only the first set of estimators suggest a sustainable positive Sharpe ratio which is an indicator for the profitability of active participation in the stock market.

Yet, if predictability in the asset returns is considered, this conclusion may change. Barberis (2000, p.243-p.245) and Campbell and Viceira (2002), among others, note that the conditional variance of cumulative stock returns may grow slower than linearly with the investor's horizon. The economic intuition follows from the negative contemporaneous correlation between dividend yield and stock returns. A decrease in dividend yield is likely to be accompanied by a contemporaneous positive shock to returns. However, since the dividend yield is lower, stock returns are forecast to be lower in the future ($\hat{\beta}_1 > 0$). The rise, followed by a fall in returns, generates a component of negative serial correlation in returns which slows the evolution of the variance of cumulative returns as the horizon grows.

A naive buy-and-hold investor who relies entirely on the estimated parameter values, interprets this effect of predictability as a reduction in risk and hence as a justification for active participation in the risky market for long- term asset allocation. Investors who account for parameter uncertainty do not necessarily invest more in the risky asset for longer horizons.

Barberis (2000) shows that an investor who adopts a Bayesian approach to uncertainty believes that conditional variances grow more quickly as the horizon grows and doubts whether the predictive power of the dividend yield is large enough to slow the evolution of conditional variances.

Also a robust investor is reluctant to interpret (estimated) predictability entirely as a reduction in risk. Based on the uncertainty of the parameters, the robust investor also considers alternative conceivable parameter configurations that may not feature predictability. As portfolio choice must be robust to such alternative parameter configurations, the robust investor may choose to not exploit the predictability featured by the parameter (point) estimates.

4.2.2 Effects of robustness and predictability

We start with an examination of robustness without engaging in actual portfolio choice and consider the (geometric) average monthly Sharpe ratio.

As predictability induces dynamics in returns over time, we calculate the average monthly Sharpe ratio for time periods ranging from one month up to 100 years. The distribution of multi-period returns follows from the return model (3)-(4) which features predictability. More specifically, we are interested in the cumulative continuously compounded excess return over T months, $y_T^a = y_1 + y_2 + .. + y_T$, which we will denote by y_T henceforth. This return and the future state are given by:

$$\begin{aligned}
\begin{pmatrix} y_T \\ x_T \end{pmatrix} \mid x_0 &\sim N\left(\begin{pmatrix} \mu_T \\ \nu_T \end{pmatrix}, \Sigma_T\right) \qquad (16)\\
\begin{pmatrix} \mu_T \\ \nu_T \end{pmatrix} &= \begin{pmatrix} T\alpha_1 + \beta_1\rho\alpha_2 + \beta_1\chi x_0 \\ \chi\alpha_2\beta_2^T x_0 \end{pmatrix}\\
\Sigma_T &= \begin{pmatrix} T\sigma_1^2 + 2\beta_1\rho\sigma_{12} + \beta_1^2\phi\sigma_2^2 & \chi\sigma_{12} + (\chi_2 - \chi)/(\beta_2 - 1)\sigma_2^2 \\ \chi\sigma_{12} + (\chi_2 - \chi)/(\beta_2 - 1)\sigma_2^2 & \chi_2\sigma_2^2 \end{pmatrix}
\end{aligned}$$

given geometric sequences

$$\begin{aligned}
\chi(T) &= \frac{\beta_2^T - 1}{\beta_2 - 1}\\
\chi_2(T) &= \frac{\beta_2^{2T} - 1}{\beta_2^2 - 1}\\
\rho(T) &= \sum_{t=1}^{T-1}(T-t)\beta_2^{t-1} = \frac{\beta_2^T - 1 + T(1-\beta_2)}{(\beta_2 - 1)^2} \qquad (17)\\
\phi(T) &= \sum_{t=1}^{T-1}\left(\frac{\beta_2^{T-t} - 1}{\beta_2 - 1}\right)^2 = \frac{T}{(\beta_2 - 1)^2} - 2\frac{\beta_2^T - 1}{(\beta_2 - 1)^3} + \frac{\beta_2^{2T} - 1}{(\beta_2^2 - 1)^3}
\end{aligned}$$

The expected log excess returns (16) depend on x_0, but the variance does not depend on x_0. Cumulative returns R_T with $1 + R_T = e^{Tr_f + y_T}$ are lognormally distributed with parameters $\mu_T + Tr_f$ and the first diagonal element, denoted σ_T^2, of Σ_T.

We consider the geometric average monthly Sharpe ratio. The geometric average cumulative T-month return is $\sqrt[T]{R_T}$ and is lognormally distributed with parameters[2] μ_T/T and σ_T^2/T^2 and, using the moments of the lognormal distribution, we derive the geometric average monthly Sharpe ratio

$$\mathrm{Sh} = \frac{e^{r_f+\mu_T/T+0.5\sigma_T^2/T^2} - e^{r_f}}{e^{r_f+\mu_T/T+0.5\sigma_T^2/T^2}\sqrt{e^{\sigma_T^2/T^2} - 1}}$$

We use $e^x \approx 1 + x$ for small x to approximate the geometric average monthly Sharpe ratio by

$$\begin{aligned} \mathrm{Sh} &= \frac{(e^{\mu_T/T+0.5\sigma_T^2/T^2} - 1)e^{r_f}}{e^{\mu_T/T+0.5\sigma_T^2/T^2}e^{r_f}\sqrt{e^{\sigma_T^2/T^2} - 1}} \\ &= \frac{1 - e^{-\mu_T/T-0.5\sigma_T^2/T^2}}{\sqrt{e^{\sigma_T^2/T^2} - 1}} \\ &\approx \frac{\mu_T}{\sigma_T} + \frac{1}{2}\sigma_T/T. \end{aligned} \tag{18}$$

If $\beta_2 < 1$, $\rho(T) = \frac{T-\chi(T)}{1-\beta_2}$ and

$$\begin{aligned} \mu_T &= T\alpha_1 + \beta_1\rho(T)\alpha_2 + (x_0 - \nu + \nu)\beta_1\chi(T) \\ &= T\alpha_1 + \beta_1\tfrac{T-\chi}{1-\beta_2}\alpha_2 + (x_0 - \nu + \tfrac{\alpha_2}{1-\beta_2})\beta_1\chi(T) \\ &= T(\alpha_1 + \tfrac{\beta_1}{1-\beta_2}\alpha_2) + (x_0 - \nu)\beta_1\chi(T). \end{aligned}$$

and

$$\sigma_T^2 = T\sigma_1^2 + 2\beta_1\rho(T)\sigma_{12} + \beta_1^2\phi(T)\sigma_2^2.$$

The uncertainty set for the robust approach is based on (11). We consider (i) a naive evaluation of the Sharpe ratio ($\theta = 0$), and (ii) a robust evaluation which implies a preference for robustness of 95% ($\theta \approx 3.1$). Moreover we add an economic consideration to the definition of the uncertainty set which asserts mean reversion for dividend yields. In particular, we impose that the long term mean $\nu = \frac{\alpha_2}{1-\beta_2}$ is finite. This is certainly the case if we append the constraint $\beta_2 \leq 1 - \epsilon$ ($\epsilon = 0.01$) to the uncertainty set,

$$\mathcal{U} \to \mathcal{U} \cap \{\xi : \beta_2 < 1 - \epsilon\} \tag{19}$$

We also define a finite approximation $\mathcal{U}_f$ to the uncertainty set. The set $\mathcal{U}_f$ is a grid of parameter values centered at $\hat{\xi}$ and with grid distances $\alpha\hat{\sigma}_\xi$ with the scalar α set such that P grid points are contained in the original set $\mathcal{U}$. We will use this finite

[2] If $R_T \sim LOGN(\mu_T, \sigma_T^2)$, then $\sqrt[T]{R_T} \sim LOGN(\mu_T/T, \sigma_T^2/T^2)$.

approximation to enhance precision in our computations.

Figure 4.2 depicts the uncertainty set for each combination of parameters. The strong correlations between the parameters cause 'narrow' ellipsoidal forms; large values for α_1 comply with small values for β_1 and α_2 and large values for β_2 and vice versa. Observe that the restriction $\beta_2 < 1$ also implies a truncation of the other parameters due to the relations associated with the ellipsoid. Also observe that the uncertainty set for the 1985-1995 dataset contains the return model without predictability $\beta_1 = 0$.

The robust average monthly Sharpe ratio is defined by

$$\min_{\xi \in \mathcal{U}} \mathrm{Sh}(\xi) \tag{20}$$

with $\mathrm{Sh}(\xi) = \frac{\mu_T(\xi)}{\sigma_T(\xi)} + \frac{1}{2}\sigma_T(\xi)/T$. The robust Sharpe ratio as defined by (20) may take negative values.

The Sharpe ratio has an immediate implication for portfolio choice. The robust Sharpe ratio serves as an upperbound on the robust mean-variance utility function $Q(w, \xi) = \mu_T(\xi, w) - \frac{1}{2}\gamma\sigma_T^2(\xi, w)$ where γ denotes the investor's risk aversion, $\mu_T(\xi, w)$ is portfolio return and $\sigma_T(\xi, w)$ is the portfolio variance. Indeed,

$$\max_{w} \min_{\xi \in \mathcal{U}} f(w, \xi) \leq \min_{\xi \in \mathcal{U}} \max_{w} f(w, \xi) = \min_{\xi \in \mathcal{U}} \frac{1}{2\gamma} \mathrm{Sh}(\xi).$$

Hence for mean-variance portfolio choice with multiple risky assets, a positive robust Sharpe ratio is not indisputable evidence for existence of a robust asset allocation with positive expected performance.

Problem (20) is a non-convex optimization problem. Fortunately it has only four variables and we solve it with a conjugate gradient method as provided by Matlab version 6.1 (R12.1) with multiple starting solutions given by the elements of $\mathcal{U}_f$ to enhance global optimality. In theory, the non-convex inner optimization problem may produce local optima for the worst case parameter value which leads to a decline in robustness of the solution. Numerical experiments[3] do not indicate that this is a problem.

Results

Figure 4.1 shows the robust average monthly Sharpe ratio for alternative investment horizons T as a function of the initial dividend yield. A naive evaluation without uncertainty ($\theta = 0$ and $\xi = \hat{\xi}$) shows that the Sharpe ratio (18) grows linearly with the initial dividend yield (see (19)), with a slope that decreases with the investment horizon. The decrease in sensitivity to the initial dividend state for long horizons is due to the mean

[3]The computed solution to the inner minimization problem does not improve when we increase the number of starting solutions to $P = 2.500$.

reversion property of dividend yields implied by the parameter values $\hat{\alpha}_2$ and $\hat{\beta}_2$. An initial dividend yield x_0 above the long term mean dividend yield ν is likely to be followed by relatively high dividend yields which cause, on the short term, larger cumulative returns. On the other hand, long-term cumulative returns crucially depend on average long-term dividend yield which converges to its long-term mean and is insensitive to the initial dividend yield.

Consider the relation between the Sharpe ratio and the investment horizon when the initial dividend yield equals the long term average ($x_0 = \nu$). The effect of predictability is visible from the increase in the Sharpe ratio when the investment horizon grows: the variance of cumulative log stock returns grows slower than linearly with the investment horizon, while cumulative log stock return grows linearly with the horizon.

The right panels of figure 4.1 depict the robust Sharpe ratio. Naturally the robust Sharpe ratio is smaller than the naive Sharpe ratio for equal initial dividend yields and investment horizons. The 1952-1995 dataset implies the smallest parameter uncertainty and consequently implies a smaller uncertainty set. It is important to observe that this uncertainty set contains exclusively $\beta_1 > 0$ and α_2, $\beta_2 > 0$. This implies that predictability is a robust property of the return model, albeit with a decreased impact, and the long term average dividend yield $\nu = \alpha_2/(1 - \beta_2)$ is positive. This leads to positive robust Sharpe ratios for cases with an initial dividend yield above 3.6% or a large investment horizon.

Figure 4.2 shows the worst case parameter configurations for alternative initial dividend yields and investment horizons, and helps to explain the behavior of the robust Sharpe ratio. Although the precise choice of the worst case parameter configurations follows from minimizing the Sharpe ratio over the set of conceivable parameter configurations $\mathcal{U}$, we observe some structural results. For a one-year investment horizon and small dividend yield, the worst case parameter configuration concentrates on small α_1 to reduce the return and small β_2 to keep a small dividend yield for the short-term horizon. Due to the high correlations between the parameters, small α_1 and β_2 imply large β_1 and α_2. The effect of alternative β_1's is less pronounced when the dividend yields are small ($x_0 = 2\%$) and α_2 is (reliably) small comparated to x_0. For large initial dividend yields ($x_0 = 5\%$) which are likely to be followed by large dividend yields on the short term, the value of β_1 becomes more important. Hence for large initial dividend yields, the worst case β_1 is reduced accompanied by a decrease in β_2. For extremely long horizons, e.g. 100 years, the initial dividend yield hardly affects the average log returns, and the worst case parameter configuration aims to minimize the long-term average dividend yield $\nu = \alpha_2/(1 - \beta_2)$ and the extent of predictability of returns β_1. The worst case parameter configurations for a 5-year investment horizon combine the aforementioned effects: When the initial dividend yield is small ($x_0 = 0.02$), the worst case parameter configurations concentrate on small α_1 and β_2. When the initial dividend yield increases to $x_0 = 0.05$, β_1 is swiftly reduced.

The 1985-1995 dataset implies larger estimation uncertainty and 'return predictability' is not robust: the uncertainty set contains the return model without predictability ($\beta_1 = 0$) and even a return model with a negative dependence on the state variable ($\beta_1 < 0$ and α_2, $\beta_2 > 0$). This leads to a negative robust Sharpe ratio irrespective of the initial dividend yield or investment horizon. For short term horizons, the worst case parameter configurations follow a similar pattern as for the 1952-1995 dataset. But as the larger uncertainty set contains negative β_1 the robust Sharpe ratio remains negative for large initial dividend yields. Moreover as the uncertainty set contains negative α_2 and consequently negative long term average dividend yields ν, the robust Sharpe ratio for long-term horizons may be negative for positive β_1. In this case a robust evaluation of the Sharpe ratio does not prefer long-term over short-term investment.

4.3 Buy-and-hold portfolio choice

A buy-and-hold investor with a constant relative risk aversion power utility function over terminal wealth of the form (1) chooses a fraction w of wealth for investment in the risky asset at the beginning ($t = 0$) of the investment horizon and maximizes the conditional expectation,

$$\mathsf{E}_0\left[\frac{W_T^{1-\gamma}}{1-\gamma} \mid \xi\right]. \tag{21}$$

Henceforth we assume that the investor's risk aversion parameter $\gamma > 1$.

We consider the relation between the optimal portfolio choice and the initial value of the predictor variable x_0, the investment horizon T, preference for robustness θ and the corresponding worst case parameter configuration.

Unfortunately a power utility function (1) prevents a simplification of the expectation of utility as a function of the first moments of the return and portfolio choice like mean-variance portfolio choice. Instead we maximize (21) numerically.

For given ξ, we approximate the conditional expectation by a finite approximation,

$$\sum_{j=1}^{J} p_j(\xi)\frac{W_T(w, y_{T,j}(\xi))^{1-\gamma}}{1-\gamma} \tag{22}$$

with

$$W_T(w, y_{T,j}(\xi)) = \left(we^{Tr_f + y_{T,j}(\xi)} + (1-w)e^{Tr_f}\right) W_0.$$

and $\{p_j(\xi), y_{T,j}(\xi)\}_{j=1}^{J}$ is a stratified sample conditional on the parameter ξ. We stratify a sample of residuals ε from the standard normal distribution in $J = 1000$ equispaced intervals with means ε_j and cumulative probabilities p_j. Conditional on x_0, we compute $y_{T,j}(\xi) = \mu_T(\xi) + \sigma_T(\xi)\varepsilon_j$ with $\mu_T(\xi)$ and $\sigma_T^2(\xi)$ given by (16) and the first diagonal element of Σ_T respectively.

For alternative values of T and ξ we use the same stratified sample of residuals. A fixed approximate distribution, even if it deviates from the original distribution, improves the comparison among alternative horizons and parameter values. Moreover, the stratified sample leads to the same results (see figures (4.4) and (4.6)) as reported by Barberis (2000) who shows that his approximation is accurate. The stratified sample is computationally faster than a large random sample, which is needed to calculate robust solutions[4].

A naive investor assumes $\xi = \hat{\xi}$ and maximizes (22). A Bayesian investor prepares an unconditional approximate return distribution. She samples $\tilde{K} = 1.000.000$ parameter values from the posterior parameter distribution and for each sampled parameter $\tilde{\xi}_{\tilde{j}}$, $\tilde{j} = 1, .., \tilde{J}$, she draws a random residual $\tilde{\varepsilon}$ from the stratified sample $\{p_k(\xi), y_{T,j}(\xi)\}_{j=1}^{J}$ and computes the corresponding conditional excess return $y_{T,\tilde{j}} = \mu_T(\tilde{\xi}_{\tilde{j}}) + \sigma_T(\tilde{\xi}_{\tilde{j}})\tilde{\varepsilon}$. The Bayesian investor solves

$$\sum_{\tilde{j}=1}^{\tilde{J}} \frac{1}{\tilde{J}} \frac{W_T(w, y_{T,\tilde{j}})^{1-\gamma}}{1-\gamma}. \tag{23}$$

The robust portfolio choice problem for a buy-and-hold investor is a maximin optimization problem,

$$\max_{w \in \mathcal{W}} \min_{\xi \in \mathcal{U}} \sum_{j=1}^{J} p_j(\xi) \frac{W_{T,j}(\xi)^{1-\gamma}}{1-\gamma}. \tag{24}$$

The objective function (22) is convex on $w \in [0, 1]$ for each parameter ξ and consequently the portfolio choice problems for a naive and a Bayesian investor are convex optimization problems on $w \in [0, 1]$ and can be solved to (global) optimality. Also the minimum of a set of convex functions is a convex function; hence the outer maximization (24) is a convex optimization problem. However, each function evaluation of the maximization problem is computationally costly as it involves solving the inner minimization problem. Moreover, the minimization problem is not a convex optimization problem on the domain $\xi \in \mathcal{U}$.

We adopt sequential optimization to solve iteratively the outer maximization of (24) with iterations that involve solving the inner minimization problem. Convergence to a global optimal solution of the non-convex inner minimization problem is enhanced by considering multiple starting solutions, given by the set $\mathcal{U}_f$ with $P = 120$ as explained in the previous section.

Results

For our computations we use a risk aversion $\gamma = 10$ so that we can compare the results to Barberis (2000). Figures 4.4 and 4.6 show the optimal portfolio allocations. The results for the naive and Bayesian investor replicate Barberis (2000).

[4]The computations for figure 4.7 take 72 hours CPU time on an AMD Athlon 2400 Mhz XP+/1GB RAM running Matlab version 6.1.0.450 (R12.1) under Debian BGU/Linux 3.0.

A naive investor increases risky asset allocation for longer horizons and higher dividend yields. In the absence of estimation uncertainty, predictability makes stock returns less risky at long horizons and optimal risky investment increases. The increase in asset allocation for above average initial dividend yields declines with the horizon as the mean reversion effect of dividend yields enters.

Barberis (2000) explains that a Bayesian approach to estimation uncertainty implies an increase in risk as the conditional variances of the stock return increase due to uncertainty in the mean stock return and uncertainty about the predictive power of the dividend yield. Consequently the risky asset allocation is smaller. The optimal investment, which is not a monotonic function of the investment horizon or initial dividend yield, results from an evaluation of predictability which makes stocks look less risky at long horizons and estimation uncertainty which makes stocks look more risky. Estimation uncertainty makes long-term portfolio allocation less sensitive to the initial dividend yield than short term portfolio allocation. The effect of uncertainty may even reduce long-term portfolio allocation below short term portfolio allocation.

A robust approach to uncertainty leads to (strikingly) different results. Consider the 1952-1995 dataset with smallest uncertainty. The risky asset allocation increases, albeit with reduced slope, monotonically over the investment horizon and initial dividend yield. As the uncertainty set contains parameter configurations which almost reduce model (3)-(4) to a static model without return predictability ($\beta_1 = 0$), the extent of predictability evaluated by a robust investor is reduced considerably as β_1 may be as small as 0.02. Moreover the long-term average dividend yield ν is positive as α_2 is reliably positive and $\beta_2 < 1$. Consequently the risky asset allocations for alternative dividend yields converge as the investment horizon increases. The robust investment is monotonic in the dividend yields as β_1 and β_2 are reliably positive. The worst case parameter configurations for the robust buy-and-hold portfolio allocation depicted in figure 4.5 follow a similar pattern as described in the previous section. For short term investment and low initial dividend yields, the worst case parameter configuration features small α_1 and α_2. For above average dividend yields, the 'predictability parameter' β_1 is reduced. For long term investment the average dividend yield converges to ν and the worst parameter configuration features small β_1.

Robust portfolio investment based on the 1985-1995 dataset with large uncertainty is zero. The associated uncertainty set contains negative α_1, β_1 and α_2. Negative α_1 lead to negative returns on the short term horizons and low dividend yields and negative β_1 cause negative returns for short term horizons with high initial dividend yields and for long term horizons (with $\alpha_2 > 0$ and $\beta_2 \approx 0.99$).

The Bayesian as well as the robust approach accounts for uncertainty, yet lead to fundamentally different strategies over time. A possible explanation follows from the slightly different priors for the Bayesian and robust investors. The robust investor, unlike the Bayesian investor, takes the prior information $\beta_2 < 1$ into account. We may improve the basis for comparison between a robust and Bayesian approach by assuming identical prior information in the two approaches. Figure 4.10 shows the optimal portfolio allocations for a Bayesian investor who does take this information into account. A Bayesian still invests more in the risky assets than a robust investor but the portfolios show similar development over time horizons. It also shows that the optimal decision of a Bayesian investor is sensitive to the prior.

4.4 Dynamic portfolio choice

In this section we consider the portfolio choice of an investor who may revise her portfolio every year to adapt to newly obtained information, in this case the dividend yield, at the end of each year.

Assume the investment horizon is K years and the investor may revise her portfolio at the beginning of each year $k = \{0, 1, .., K-1\}$ whereupon the portfolio is invested for the period $[t_k, t_{k+1})$. The value of the state variable dividend yield at the beginning of year k is denoted by x_k and the cumulative continuously compounded excess return over the period $[t_k, t_{k+1})$ conditional on x_k is denoted y_{k+1,x_k}. A dynamic portfolio allocation w is a plan that indicates the investment w_{k,x_k} for each decision moment t_k, $k = 0, .., K-1$ and concurrent dividend yield x_k. We denote wealth accumulated at t_k in state x_k by W_{k,x_k}. Moreover, we use W_K to refer to the wealth at the end of the horizon.

Conditional on the parameter ξ, which we omit in the notation for now, the dynamic portfolio choice problem is

$$\max_{w \in \mathcal{W}} \mathsf{E}_0 \left[\frac{{W_K}^{1-\gamma}}{1-\gamma} \right]. \tag{25}$$

with wealth transitions

$$\begin{aligned} W_{k+1,x_{k+1}} &= (1 + R_{k+1,x_k}(w))\, W_{k,x_k} \\ W_K &= \left(1 + R_{K,x_{K-1}}(w)\right) W_{K-1,x_{K-1}}. \end{aligned}$$

and portfolio return

$$1 + R_{k+1,x_k}(w) = (w_{k,x_k} e^{r_f + y_{k+1,x_k}} + (1 - w_{k,x_k}) e^{r_f}).$$

and $\mathcal{W}$ presents the set of portfolios w with $w_{k,x_k} \in [0, 1]$ for all t_k and x_k. Note that the expectation operator in (25) is a function of the returns y_{k,x_k} and the transition probabilities between states $[x_k, x_{k+1}]$ in subsequent time periods $[t_k, t_{k+1})$. Problem (25) presents a dynamic optimization problem suitable for applications of the Bellman optimality principle. The derived utility of wealth at any state (t_k, x_k) is

$$J(W_{k,x_k}, t_k, x_k) = \max_{w} \mathsf{E}_{t_k} \left[\frac{W_K^{1-\gamma}}{1-\gamma} \mid W_{k,x_k}, t_k, x_k \right] \tag{26}$$

which is the maximal expected (final) utility as expected at time t_k in state x_k. The

Bellman principle of optimality states

$$J(W_{k,x_k}, t_k, x_k) = \max_{w_{k,x_k}} \mathsf{E}_{t_k} \left[J(W_{k+1,x_{k+1}}, t_{k+1}, x_{k+1}) \mid W_{k,x_k}, t_k, x_k \right]. \tag{27}$$

For a more convenient expression, define

$$Q(t_k, x_k) = \max_{w_{k,x_k}} \mathsf{E}_{t_k} \left[\frac{W_K^{1-\gamma}}{W_{k,x_k}^{1-\gamma}} \mid t_k, x_k \right] \tag{28}$$

and note that (by homogeneity in W of degree $1-\gamma$),

$$J(W_{k,x_k}, t_k, x_k) = \frac{W_{k,x_k}^{1-\gamma}}{1-\gamma} Q(t_k, x_k). \tag{29}$$

Substitution of (29) in the Bellman equation (27),

$$Q(t_k, x_k) = \max_{w_{k,x_k} \in [0,1]} \mathsf{E}_{t_k} \left[(1 + R_{k+1,x_k}(w_{k,x_k}))^{1-\gamma} Q(t_{k+1}, x_{k+1}) \middle| t_k, x_k \right]. \tag{30}$$

Note that (28) implies $Q(t_K, \cdot) = 1$ and, due to the investor's constant relative risk aversion, W does not enter expression (30)[5]. Given $Q(0, x_0)$ and initial wealth W_0, the expected terminal utility of portfolio choice w is $W_0/Q(0, x_0)$. Hence solving (25) is equivalent to computing $Q(0, x_0)$. Moreover, if we follow Barberis (2000) and adopt a finite state state space for the dividend yield, we can solve for $Q(0, x_0)$ by backward induction using (30).

We follow Barberis (2000) and discretize the state space in 25 equally spaced grid points on the interval ranging from three standard deviations below the historical mean to three standard deviations above.

To calculate the expectation in (30) we rely on a sample $\{p_j(\xi), x_{k+1,x_k,j}, y_{k+1,x_k,j}(\xi)\}_{j=1}^J$ based on a stratified sample of two-variate residuals $\varepsilon \sim N(0, I_2)$ in $J = 1000$ equi-spaced intervals with means ε_j and cumulative returns p_j. Conditional on x_k, we compute

$$\begin{pmatrix} y_{k+1,x_k,j}(\xi) \\ x_{k+1,j}(\xi) \end{pmatrix} = \begin{pmatrix} \mu_T(\xi) \\ \nu_T(\xi) \end{pmatrix} + A_T(\xi)\varepsilon_j \tag{31}$$

with $A_T(\xi)$ the lower triangular part of the Choleski decomposition of $\Sigma_T(\xi)$ and $(\mu_T(\xi), \nu_T(\xi))$ and $\Sigma_T(\xi)$ given by (16). As we assume yearly portfolio revisions, we consider $T = 12$-monthly returns. To fit the finite state space, we match each $x_{k+1,x_k,j}(\xi)$ to the closest dividend yield in the discretized state space.

[5]Without constant relative risk aversion, W is a state variable and enters the state space definition. The size of an accurate approximation of this multi-dimensional state space grows exponentially with the number of state variables and complicates the computations.

$$Q(t_k, x_k) = \max_{w_{k,x_k} \in [0,1]} \sum_{j=1}^{J} p_j(\xi) \left((1 + R_{k+1,x_k,j}(w_{k,x_k}, \xi))^{1-\gamma} Q(t_{k+1}, x_{k+1,x_k,j}) \right). \quad (32)$$

with

$$R_{k+1,x_k,j}(w, \xi) = (w_{k,x_k} e^{r_f + y_{k+1,x_k,j}(\xi)} + (1 - w_{k,x_k}) e^{r_f}) - 1.$$

We keep the stratified sample of residuals constant to fix the return distribution, albeit an approximation to the original distribution, to improve the comparison among alternative horizons, parameter values and investors.

The naive investor solves the backward recursion (32) conditional on $\xi = \hat{\xi}$. A Bayesian investor samples from the unconditional approximate return distribution as described in the previous section.

A robust investor solves

$$\max_{w \in \mathcal{W}} \min_{\xi \in \mathcal{U}} \mathsf{E}_0 \left[\frac{{W_K}^{1-\gamma}}{1-\gamma} | \xi \right] \quad (33)$$

To calculate the expectation, we rely on the stratified sample. In this case, a straightforward implementation of (33) demands the construction of scenario paths for each evaluated parameter configuration ξ. As the scenarios paths recombine, a backward recursion is a fast way to evaluate the objective of the inner minimization problem for w given.

Therefore we define

$$Q(t_k, x_k, w, \xi) = \sum_{j=1}^{J} p_j(\xi) \left((1 + R_{k+1,x_k,j}(w, \xi))^{1-\gamma} Q(t_{k+1}, x_{k+1,j}(\xi), w, \xi) | \, x_k, \xi \right). \quad (34)$$

Given a portfolio choice w and a parameter ξ, $W_0/Q(0, x_0, w, \xi)$ is equal to the objective value $\mathsf{E}_0[\frac{W_K^{1-\gamma}}{1-\gamma} | \, \xi]$ associated with portfolio choice w. Therefore problem (33) is equivalent to

$$\max_{w \in \mathcal{W}} \min_{\xi \in \mathcal{U}} W_0/Q(0, x_0, w, \xi). \quad (35)$$

Observe that for given w and ξ, $Q(0, x_0, w, \xi)$ is readily computed by (34).

We use sequential optimization as explained in the previous section to solve (33) to reasonable accuracy.

Results

The optimal initial investment in the risky asset of a naive investor who rebalances once a year, increases when the investment horizon is longer or initial dividend yield is larger. Barberis (2000) explains this by hedging demands described by Merton (1973), which arise from the negative correlation between the expected return and dividend yield on the one side and realized return on the other side. Increases in the dividend yield are likely to decrease the realized return but also lead to an increase in the expected return. Hence risky investment provides higher wealth precisely when investment opportunities worsen.

The initial investment of a naive dynamic investor is not necessarily larger than the investment of a buy-and-hold investor for small initial dividend yields as the dynamic optimal solution awaits more prosperous times for active investment.

A Bayesian approach to estimation uncertainty, which implies an increase in the conditional variance and obscures the predictability of returns, leads to slightly lower optimal investment in the risky asset.

A robust approach to estimation uncertainty leads to considerably lower investment than naive optimal investment and reduces to zero when uncertainty is large (1985-1995 dataset). Whereas Bayesian investors have widely fluctuating investments over time as the dividend yield evolves, the robust investment is stable.

As the 1952-1995 dataset implies predictability for all conceivable parameters, albeit in less extent, the robust asset allocation is larger for long horizons and higher dividend yields. However differences in investment among horizons are reduced considerably because the extent of predictability (β_1) is small. The doubt about predictability also leads to similarity of robust buy-and-hold and dynamic portfolio allocations.

Figure 4.8 shows the worst case parameter configurations for various investment horizons and alternative initial dividend yields which have a similar pattern as the worst case parameter configurations for buy-and-hold strategies.

4.5 Concluding remarks and further research

A robust approach to parameter uncertainty leads to decisions which are quite different from a naive or even a Bayesian approach, both for buy-and-hold as well as dynamic portfolio choice.

The robust investor is only active when the investment opportunities are, even if there is parameter uncertainty, significant. This implies active investment when uncertainty is small, which is the case for the 52-95 dataset, but when predictability is uncertain and future returns are not significantly different from zero (as for the 85-95 dataset), the robust investor remains passive.

It is interesting to consider the parameter configurations which are worst case for the robust decision. The return model, which is a bi-linear function of the uncertain parameters and the relations between conceivable parameter values implied by the uncertainty set, make the determination of the worst case parameter configuration a non-trivial task. Nevertheless the worst case parameter configurations show some structure.

The worst case parameter configurations tell which aspects of the return model are important for portfolio choice with different investment horizons and alternative initial states. For example, the worst case parameter configuration will only feature small long term dividend yield if this is a crucial determinant of the investment opportunity set for that particular investment horizon.

For short-term investment and a small initial dividend-yield, the worst case parameter configuration features small α_1 which directly reduces short term returns, and small β_2 which makes the small initial dividend yield persistent. For large initial dividend yields which are likely to be followed by large dividend yields on the short term, the value of β_1 is more important and consequently reduced. For long-term investment, the long-term average dividend yield $\nu = \alpha_2/(1-\beta_2)$ is crucial. For the 52-95 dataset, ν and the predictability parameter β_1 are maximally reduced in the worst case parameter configuration.

Barberis (2000) also considers learning in the presence of estimation uncertainty. However learning about the true parameter value is incompatible with a robust approach to (parameter) estimation uncertainty. A dynamic investment strategy which adapts to learning of the uncertain parameters should include the investor's belief about the parameters as state variables. The investor's belief is typically a function of the newly realized returns which depend on the true parameters. Hence a portfolio which is robust to parameter uncertainty it must also be robust to learning. References for learning are Brennan (1998), Xia (2001) and Kandel and Stambaugh (1996).

In this study we confined to a description of the optimal portfolio allocation. An interesting direction for future research would be to perform an empirical study on the performance of naive, Bayesian and robust multi-period portfolio choice. Moreover we study robust portfolio choice based on a specific return model and utility function. Although the basic results presumably hold for alternative return models which are based on other state variables (see Barberis (2000)) and other utility functions that are based on a risk-return tradeoff, generalizations to multi-period portfolio choice with multiple assets and utility functions that depend on intermediate consumption are interesting to consider. Such problems will also be computationally more challenging.

A.4 Figures and tables

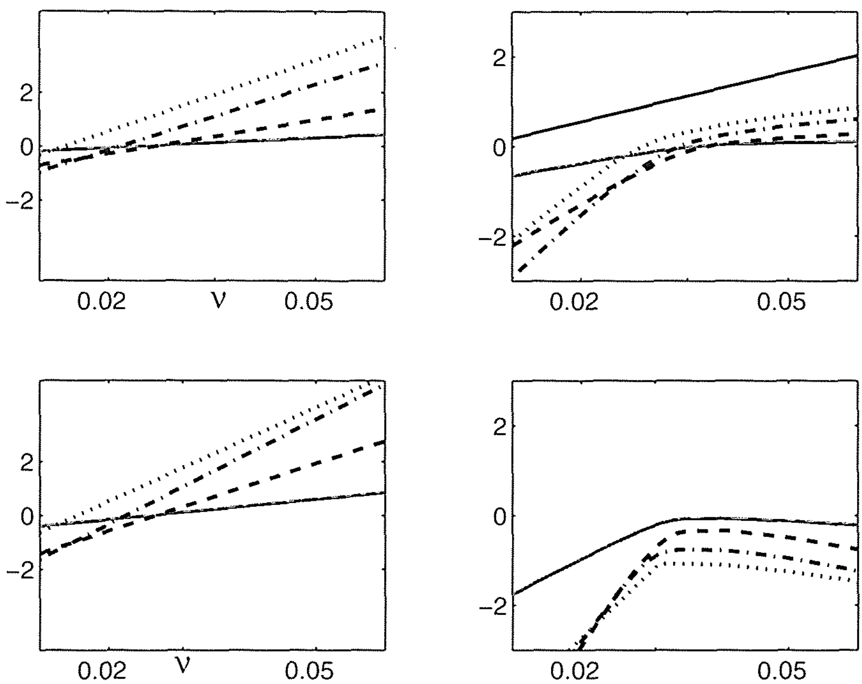
Table 4.1: Parameter estimates for a VAR model of stock returns

1952-1995		1985-1995	
μ	σ^2	μ	σ^2
0.0050	0.0017	0.0065	0.0019
(0.0018)	(0.0001)	(0.0039)	(0.0003)
α	β	α	β
-0.0143	0.5118	-0.0303	1.0919
(0.0081)	(0.2129)	(0.0281)	(0.8265)
0.0008	0.9774	0.0013	0.9577
(0.0003)	(0.0091)	(0.0010)	(0.0305)
Σ		Σ	
0.0017	**−0.9351**	0.0019	**−0.9323**
(0.0001)	(0.0055)	(0.0003)	(0.0122)
	3.0E-6		2.6E-6
	(1.9E-7)		(3.4E-7)

	α_1	β_1	α_2	β_2	α_1	β_1	α_2	β_2
α_1	0.0081	**−0.98**	**−0.94**	**0.91**	0.0281	**−0.99**	**−0.93**	**0.92**
β_1		0.2129	**0.91**	**−0.94**		0.8265	**0.92**	**−0.93**
α_2			0.0003	**−0.98**			0.0010	**−0.99**
β_2				0.0091				0.0305

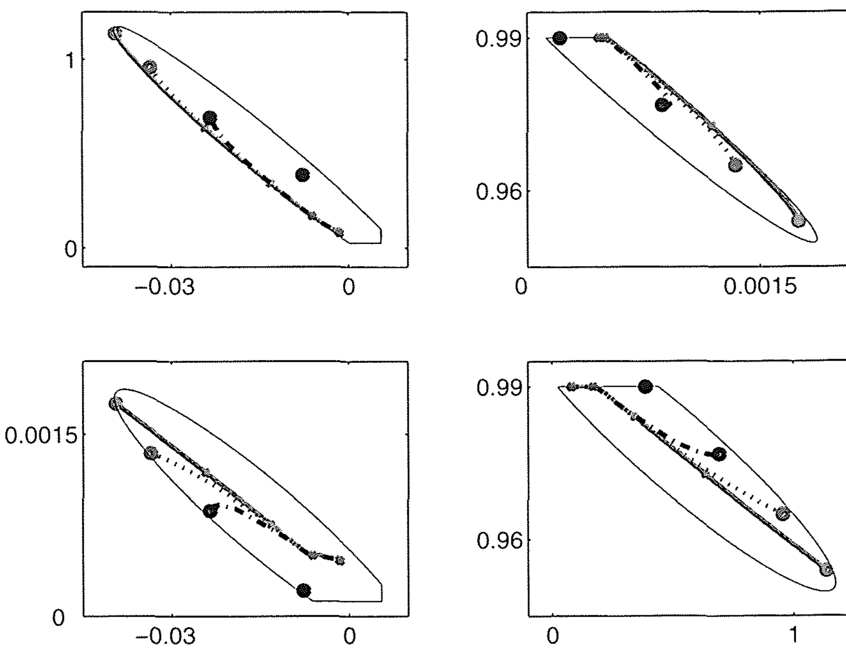
Source Barberis (2000). The results in the upper panel of this table are based on the model $r_t = \mu + \varepsilon_t$, where r_t is the continuously compounded excess stock index return in month t and $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. The results in the middle panels are based on the return model (3)-(4) with dividend yield as predictor variable. The table gives the mean and, in parentheses, standard error of each parameter's posterior distribution. The lower panel describes the parameter uncertainty by the uncertainty matrix Ω corresponding to (11). The values in bold above the diagonal in the variance matrix denote correlations.
The left panel uses the 523 observations from June 1952 to December 1995; the right panel uses 120 observations from January 1985 to December 1995. The values in bold above the diagonal in the variance matrix denote correlations.

Figure 4.1: Sharpe ratio as a function of the initial dividend yield



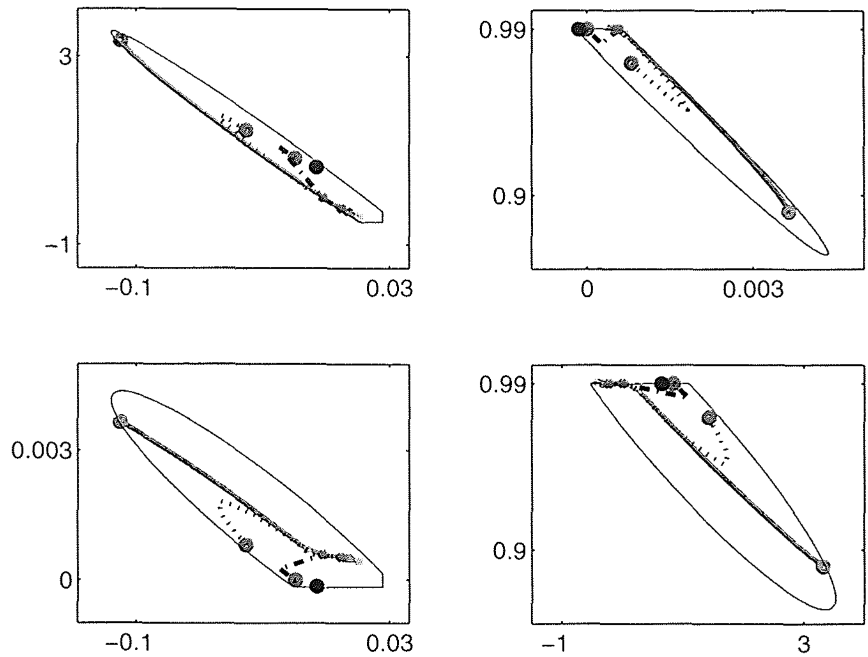
Notes: The figure describes the average monthly Sharpe ratio as a function of the initial dividend yield. The results are based on return model (3)-(4). The upper and lower panels correspond to the 1952-1995 and 1985-1995 dataset respectively. The left and right panels correspond to a naive and robust ($\theta = 3.1$) evaluation of the Sharpe ratio. The curves within each graph correspond to different investment horizons T: $T = 1$ (solid grey), $T = 12$ (dashed), $T = 60$ (dash-dot), $T = 120$ (dotted). The curve for an extremely long horizon $T = 1200$ months (solid) falls within the depicted domain for a robust evaluation on the 1952-1985 dataset.

Figure 4.2: Worst case parameter configurations



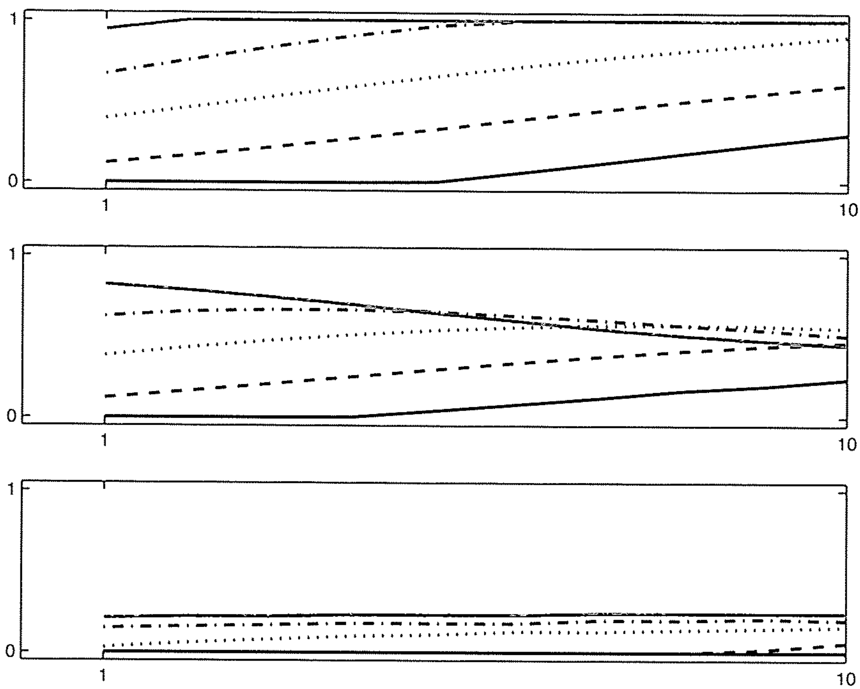
Notes: The figure shows a parametric plot of the worst parameter configurations for the average monthly Sharpe ratio. Results are based on return model (3)-(4) estimated on the 1952-1995 dataset. The panels correspond to projections of the parameter space on two dimensional subsets of parameters: (α_1, β_1) (upper-left), (α_2, β_2) (upper-right), (α_1, α_2) (lower-left) and (β_1, β_2) (lower-right). The truncated ellipsoid within each figure presents the projection of the uncertainty set (19) for $\theta = 3.1$ on the relevant parameter space. Each curve connects the worst case parameter configurations for values of the initial dividend yield between $[0.01, 0.06]$ and corresponds to a given investment horizon: $T = 12$ months (solid grey), $T = 60$ months (dotted grey), $T = 120$ months (dash-dot) and $T = 1200$ (black dot) months. The worst case parameter configuration associated with the smallest initial dividend yield is indicated by a dot.

Figure 4.3: Worst case parameter configurations



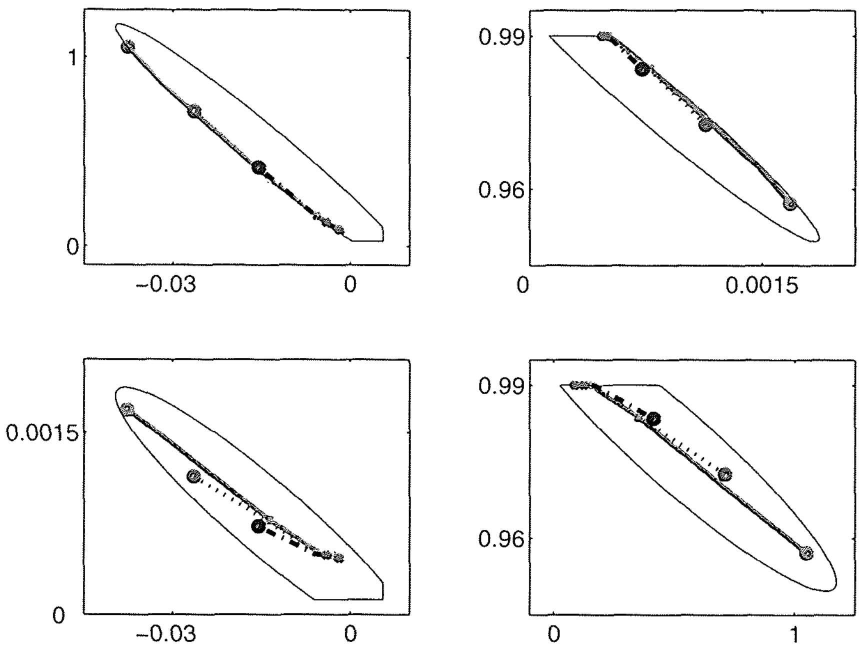
Notes: The figure shows a parametric plot of the worst parameter configurations for the average monthly Sharpe ratio. Results are based on return model (3)-(4) estimated on the 1985-1995 dataset. The panels correspond to projections of the parameter space on two dimensional subsets of parameters: (α_1, β_1) (upper-left), (α_2, β_2) (upper-right), (α_1, α_2) (lower-left) and (β_1, β_2) (lower-right). The truncated ellipsoid within each figure presents the projection of the uncertainty set (19) for $\theta = 3.1$ on the relevant parameter space. Each curve connects the worst case parameter configurations for values of the initial dividend yield between $[0.01, 0.06]$ and corresponds to a given investment horizon: $T = 12$ months (solid grey), $T = 60$ months (dotted grey), $T = 120$ months (dash-dot) and $T = 1200$ (black dot) months. The worst case parameter configuration associated with the smallest initial dividend yield is indicated by a dot.

Figure 4.4: Optimal asset allocation as a function of the investment horizon in years



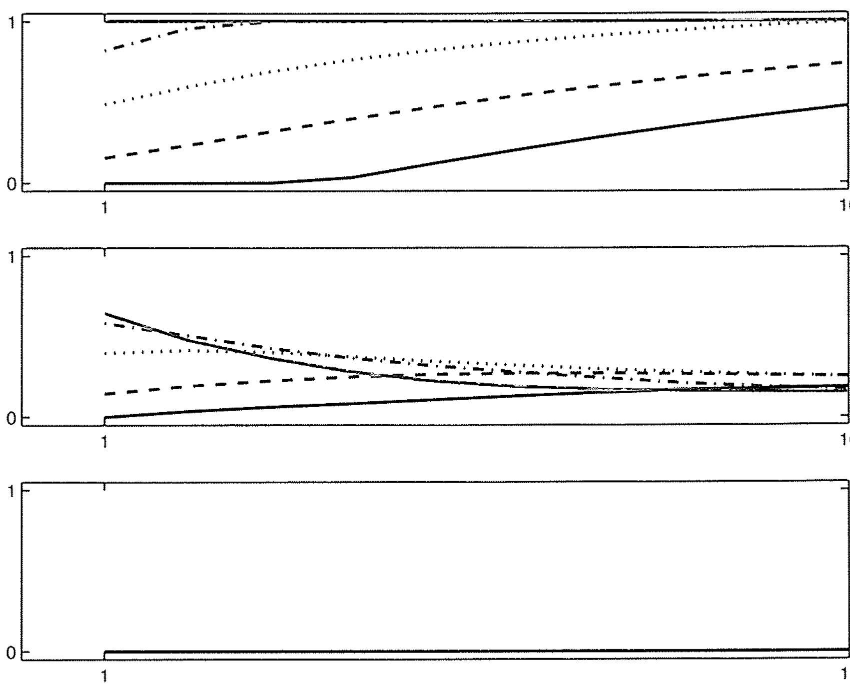
Notes: The figure describes the optimal buy-and-hold strategy of an investor who maximizes power utility function (1) over terminal wealth. The returns are estimated and predicted using (3)-(4) on the 1952-1995 dataset. The three panels describe the initial asset allocation as a function of the investment horizon in years, for a naive investor (upper panel), a Bayesian investor (middle panel) and a robust investor (lower panel). The curves within each graph correspond to alternative values of the initial dividend yield x_0: $x_0 = 2.06\%$ (solid), $x_0 = 2.91\%$ (dashed), $x_0 = 3.75\%$ (dotted), $x_0 = 4.59\%$ (dash-dot) and $x_0 = 5.43\%$ (solid grey).

Figure 4.5: Worst case parameter configuration



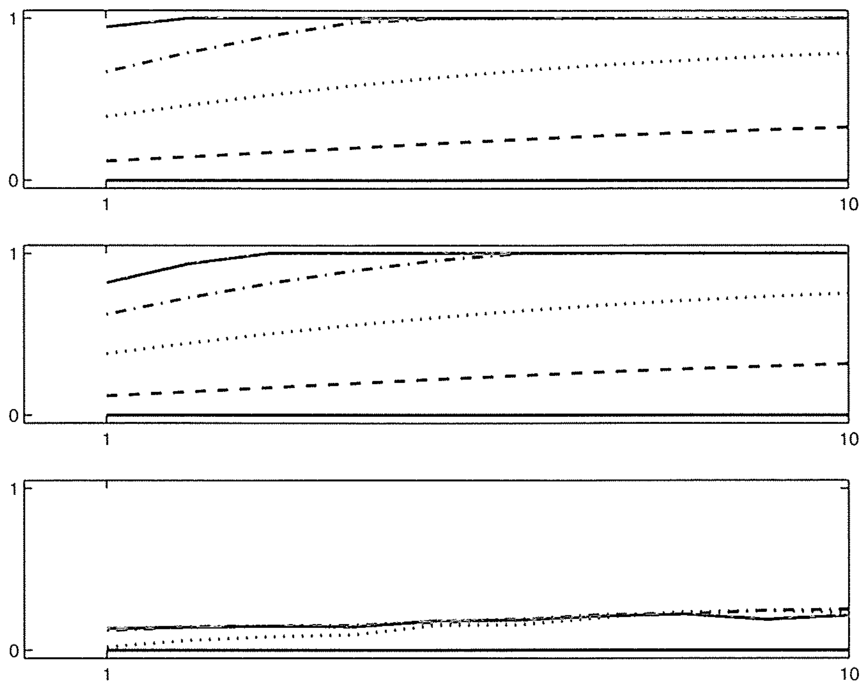
Notes: The figure describes the worst case parameter configurations for a buy-and-hold investor who maximizes power utility function (1) over terminal wealth. The returns are estimated and predicted using model (3)-(4) on the 1952-1995 dataset. The panels correspond to projections of the parameter space on two dimensional subsets of parameters: (α_1, β_1) (upper-left), (α_2, β_2) (upper-right), (α_1, α_2) (lower-left) and (β_1, β_2) (lower-right). The truncated ellipsoid within each figure presents the projection of the uncertainty set (19) for $\theta = 3.1$ on the relevant parameter space. Each curve connects the worst case parameter configurations for values of the initial dividend yield between 2.06% and 5.43% and corresponds to a given investment horizon: $T = 12$ months (solid grey), $T = 60$ months (dotted grey), $T = 120$ months (dash-dot). The worst case parameter configuration associated with the smallest initial dividend yield is indicated by a dot.

Figure 4.6: Optimal asset allocation as a function of the investment horizon in years



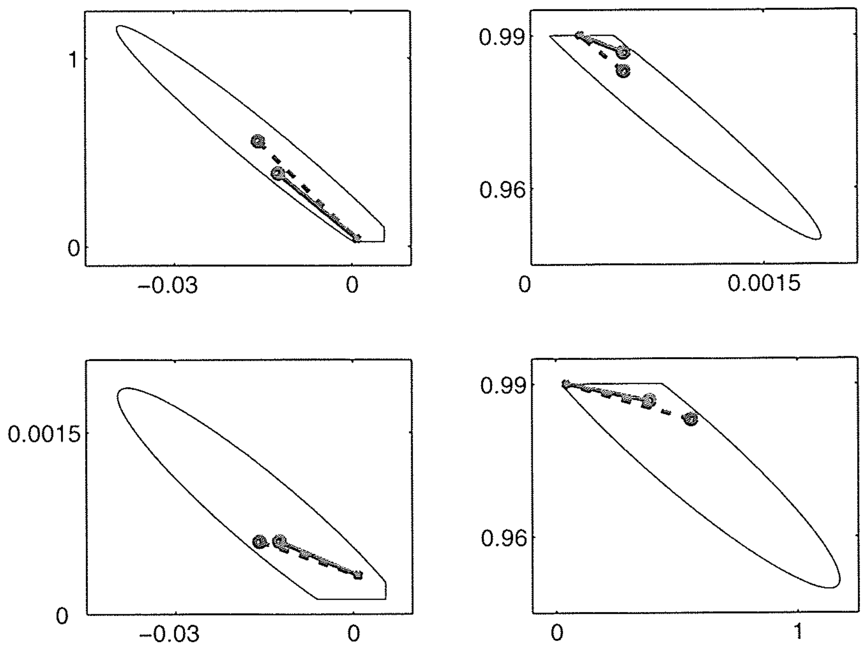
Notes: The figure describes the optimal buy–and–hold strategy of an investor who maximizes power utility function (1) over terminal wealth. The returns are estimated and predicted using model (3)-(4) on the 1985-1995 dataset. The three panels describe the initial asset allocation as a function of the investment horizon in years, for a naive investor (upper panel), a Bayesian investor (middle panel) and a robust investor (lower panel). The curves within each graph correspond to alternative values of the initial dividend yield x_0:x_0 = 2.36% (solid), x_0 = 2.86% (dashed), x_0 = 3.36% (dotted), x_0 = 3.86% (dash-dot) and x_0 = 4.36% (solid grey).

Figure 4.7: Optimal asset allocation as a function of the investment horizon in years



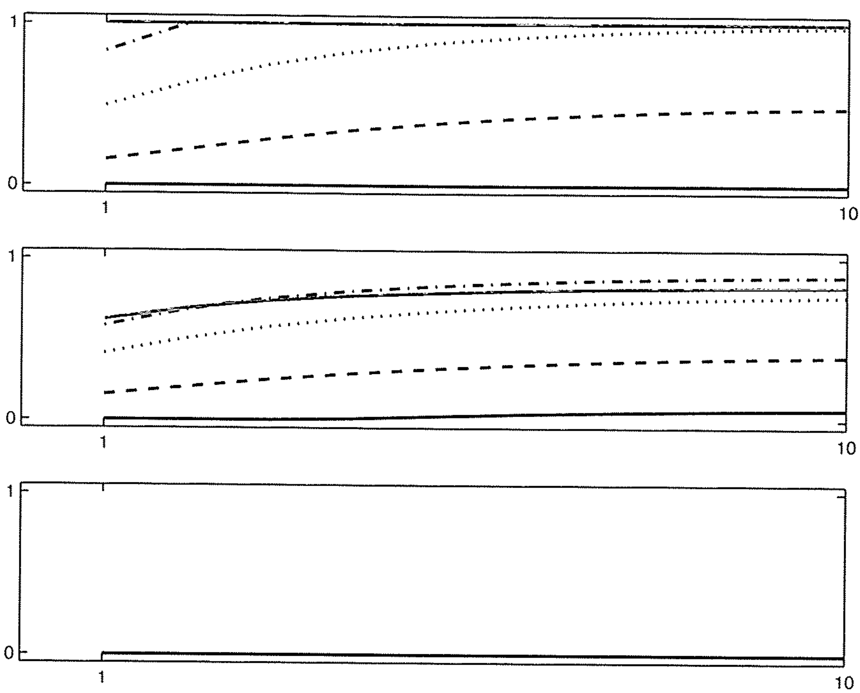
Notes: The figure shows the optimal strategy of an investor who rebalances once a year and who maximizes power utility function (1) over terminal wealth. Returns are estimated and predicted using model (3)-(4) on the 1952-1995 dataset. The three panels describe the initial asset allocation as a function of the investment horizon in years, for a naive investor (upper panel), a Bayesian investor (middle panel) and a robust investor (lower panel). The curves within each graph correspond to alternative values of the initial dividend yield x_0: $x_0 = 2.06\%$ (solid), $x_0 = 2.91\%$ (dashed), $x_0 = 3.75\%$ (dotted), $x_0 = 4.59\%$ (dash-dot) and $x_0 = 5.43\%$ (solid grey).

Figure 4.8: Worst case parameter configuration



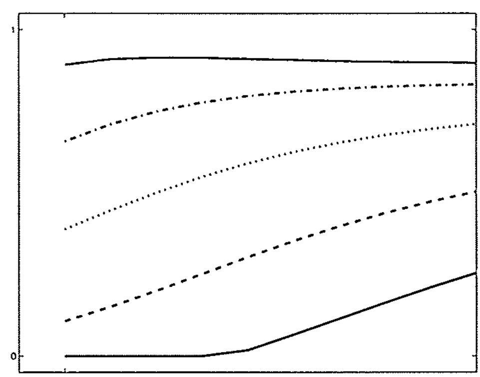
Notes: The figure describes the worst case parameter configurations for a dynamic investor who maximizes power utility function (1) over terminal wealth. The returns are estimated and predicted using model (3)-(4) on the 1952-1995 dataset. The panels correspond to projections of the parameter space on two dimensional subsets of parameters: (α_1, β_1) (upper-left), (α_2, β_2) (upper-right), (α_1, α_2) (lower-left) and (β_1, β_2) (lower-right). The truncated ellipsoid within each figure presents the projection of the uncertainty set (19) for $\theta = 3.1$ on the relevant parameter space. Each curve connects the worst case parameter configurations for values of the initial dividend yield between 2.06% and 5.43% and corresponds to a given investment horizon: $T = 60$ months (dashed grey), $T = 120$ months (solid). The worst case parameter configuration associated with the smallest initial dividend yield is indicated by a dot.

Figure 4.9: Optimal asset allocation as a function of the investment horizon in years



Notes: The figure shows the optimal strategy of an investor who rebalances once a year and who maximizes power utility function (1) over terminal wealth. Returns are estimated and predicted using model (3)-(4) on the 1985-1995 dataset. The three panels describe the initial asset allocation as a function of the investment horizon in years, for a naive investor (upper panel), a Bayesian investor (middle panel) and a robust investor (lower panel). The curves within each graph correspond to alternative values of the initial dividend yield x_0: $x_0 = 2.36\%$ (solid), $x_0 = 2.86\%$ (dashed), $x_0 = 3.36\%$ (dotted), $x_0 = 3.86\%$ (dash-dot) and $x_0 = 4.36\%$ (solid grey).

Figure 4.10: Alternative prior for the Bayesian approach



Notes: The figure shows the optimal strategy of a Bayesian buy-and-hold investor under the same assumptions as figure 4.4 but with a posterior distribution which assumes $p(\xi|\, X, Y) = 0$ if $\beta_2 \geq 0.99$.

Chapter 5

Robust Option Modelling[1]

The more options
the more difficult the choice.

In this chapter we develop the methodology to include options in a robust approach to single period portfolio choice.

An option is the right but not the obligation to buy or sell its underlying asset for a predetermined price, called the exercise price. Upon maturity, the option's value derives, as a non-negative piecewise-linear function, from its underlying stock price: depending on the stock price relative to the exercise price, the option is in-the-money and has strictly positive value or is out-of-the-money and has value zero. As a consequence, uncertainty about the underlying asset's return model affects the option return insofar as the option is in-the-money, i.e. when the option has strictly positive value. Moreover, as the option return is driven by the stock return, we may not consider uncertainty in stock and option returns separately. For example, a long position in both a stock and a put option on this stock have opposite dependence on the stock return; higher stock returns are profitable for the stock investment but have a negative effect on the put option value and vice versa. If disregarded in the robust optimization model, these issues lead to unnecessary conservatism. On the other hand, a robust model which accounts for these effects features a form of uncertainty that complicates solutions and requires new theory.

The organization of this chapter is as follows. First we give an example of a robust portfolio choice problem with one stock and one option. In section 5.2 we introduce our definitions and notation for the portfolio choice problem with multiple assets and options. In section 5.3 we consider a robust approach to this general portfolio choice problem in the presence of uncertainty in the mean returns. The main result is a transformation of the robust counterpart of a linear portfolio restriction to a constraint which can be

[1]This chapter is based on F. Lutgens and J. Sturm, 'Robust Option Modelling', CentER Working Paper, 2002.

handled by standard optimization software. We illustrate the methodology for robust option modelling as developed in section 5.3 on a benchmark tracking problem which involves options. In section 5.4 we develop the corresponding robust model and in section 5.5 we study the performance of the robust approach with a modest empirical analysis. The appendix A.5 is devoted to a duality result which enables the transformation used in section 5.3.

5.1 Example

Consider an investor who can invest in a stock with expected return μ and a call option on that stock. The investment horizon is equal to the time until expiration of the call option.

Suppose the investor wishes to choose the portfolio that maximizes the expected return. The solution is to invest all wealth in the asset with the highest expected return. If the option is fairly priced, this will usually be the call option, since this is in essence a levered investment in the stock. Consequently it will exhibit both more risk and a higher expected return. Exceptions could occur if the expected return μ is below the riskfree rate. In that case a long position in the call option could have a perverse return distribution. When the investor is not certain about the expected return on the stock, his optimal portfolio is less trivial. For some values of μ he might want to be long in the call option, for other values the highest expected return requires a short position in the call.

Suppose the investor is a robust decision maker. For each portfolio he considers the worst possible value for μ. His optimal portfolio is the best worst case. To put more structure on the uncertainty, suppose that the investor is willing to believe that μ is with certainty in the interval $\mathcal{U} = [\mu_L, \mu_U]$.

To formalize the robust portfolio problem we introduce some further notation. Let w be the fraction of wealth invested in the stock and $w^d = 1 - w$ be the fraction of wealth invested in the option. In this example we do not restrict short-selling, so both w and w_d can be negative. Maximizing robust expected return is defined as the problem

$$\max_{w} \min_{\mu \in U} w\mu + (1-w)\mu^d, \tag{1}$$

where μ^d is the expected return of the call. Since the return on an option is a nonlinear function of the return on the underlying stock, the expected return μ^d will in general also be a nonlinear function of μ. The inner minimization is therefore a nonlinear optimization problem in μ. Hence with options we cannot use equation (13) in chapter 1 to solve the problem. The solution of the inner minimization problem will depend on w.

To solve the robust problem we assume that the return distribution is discrete with a finite number K of possible outcomes. The return distribution is characterized by

$$r_k = \mu + \epsilon_k, \qquad k = 1, \ldots, K, \tag{2}$$

where each ϵ_k occurs with probability π_k. We therefore have $0 < \pi_k \leq 1$, $\sum_k \pi_k = 1$ and $\sum_k \pi_k \epsilon_k = 0$. The variance of the return distribution is $\sigma^2 = \sum_k \pi_k \epsilon_k^2$. For notational convenience the shocks are in increasing order: $\epsilon_k < \epsilon_{k+1}$.

Let S_0 be the current price of the stock, let c_0 denote the current price of the call, and let X be the strike price. The option ends in-the-money if $S_0(1 + r_k) > X$. The return on the call is therefore

$$r_k^d = \begin{cases} -1 & \text{if} \quad 1 + r_k < X/S_0 \\ ar_k + b & \text{if} \quad 1 + r_k \geq X/S_0 \end{cases} \tag{3}$$

where

$$\begin{aligned} a &= \frac{S_0}{c_0} \\ b &= \frac{S_0 - X - c_0}{c_0} \end{aligned}$$

For a given μ, let $I = \{k : 1 + r_k \geq X/S_0\}$ describe the set of return outcomes for which the call ends in-the-money, and let $\bar{I}$ be its complement. The expected return on the call option follows as

$$\begin{aligned} \mu^d &= \sum_k \pi_k r_k^d \\ &= \sum_{k \in I} \pi_k (a\mu + a\epsilon_k + b) + \sum_{k \in \bar{I}} \pi_k (-1) \end{aligned} \tag{4}$$

For some (small) k the option could be out-of-the-money for all values of $\mu \in \mathcal{U}$. For large values of k it could be in-the-money for all $\mu \in \mathcal{U}$. Finally, there might exist intermediate values of k such that the option is in-the-money for a sub-interval of $\mathcal{U}$ and out-of-the-money for the complement of $\mathcal{U}$.

As long as the set I does not change, the expected return μ^d is linear in μ. We will exploit the discreteness of the state space in the following way. Start with the lower bound $\mu_1 = \mu_L$ and consider the initial set of in-the-money returns $I_1 = \{k : 1 + \mu_1 + \epsilon_k > X/S_0\}$. Increase μ until at some point, when $\mu = \mu_2$, the largest element from $\bar{I}_1$ will be just in-the-money. Define a new set $I_2 = \{k : 1 + \mu_2 + \epsilon_k > X/S_0\}$. Continue increasing μ until the next ϵ_k switches from the out-of-the-money set to the in-the-money set or until we reach the upper bound μ_U. This leads to a partitioning of the interval $\mathcal{U}$ into J subintervals $\mathcal{U}_j$. Within each of the intervals the relation (4) between μ^d and μ is linear, while the overall relation between μ^d and μ is piecewise linear with $J - 1$ breakpoints.

With the discretization of the state space and the resulting partitioning of $\mathcal{U}$ we can solve the inner minimization problem in (1). Due to the piecewise linear relation, the minimum will be obtained at one of the breakpoints μ_j (including the endpoint $\mu_J = \mu_U$). For each w we only need to check a finite number of J points μ_j to find the minimum. We can then solve for the optimal w by linear programming.

A numerical example illustrates the procedure. Let $S_0 = 1$ and $X = 1.05$. The return distribution is approximated by $K = 5$ different outcomes. Values for π_k and ϵ_k in the numerical example are tabulated below:

k	1	2	3	4	5
π_k	0.0625	0.25	0.375	0.25	0.0625
ϵ_k	-0.2	-0.1	0	0.1	0.2

For a fair value of the call option we calculate the discounted expected value of the payoffs. Setting the riskfree rate equal to $r_f = 3\%$, the current call price is[2]

$$c_0 = \frac{1}{1+r_f} \sum_k \pi_k \max\left(0, S_0(1 + r_f + \epsilon_k) - X\right) = 0.03034$$

Assume that the uncertainty set is $\mathcal{U} = [2\%, 6\%]$. For $\mu_1 = 2\%$, the call option is in-the-money for $I_1 = \{4, 5\}$. State $k = 3$ will be just in-the-money for $\mu_2 = 5\%$. No further states change set for $\mu < \mu_U = 6\%$. This implies the partitioning $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2$ with $\mathcal{U}_1 = [2\%, 5\%)$ and $\mathcal{U}_2 = [5\%, 6\%]$. After the partitioning we can write the robust portfolio problem in the linear programming format

$$\max_{w,x} x \tag{5}$$

subject to

$$2w - 7.3(1-w) < x \tag{6a}$$
$$5w + 23.6(1-w) < x \tag{6b}$$
$$6w + 46.3(1-w) < x \tag{6c}$$

The inequalities follow directly by substituting $\mu_1 = 2\%$, $\mu_2 = 5\%$ and $\mu_3 = 6\%$ in (1) and (4). With these numerical values the first restriction has a positive coefficient on w, while the second and third have negative coefficients on w. Figure 5.1 shows the three lines and the optimal solution. The optimal portfolio is at $w = 1.08$. The associated expected return is $x = 2.75\%$. The optimal portfolio implies that the investor writes call options and invests in the underlying.

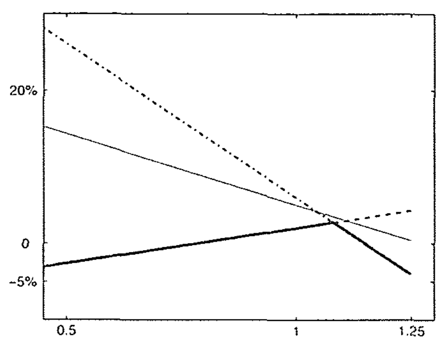
In this example a mixed portfolio is optimal because the riskfree rate r_f is inside the uncertainty interval $\mathcal{U}$. If we would set $r_f < \mu_L = 2\%$ the optimal portfolio is to go long

[2] In an actual application the price of the option will be observed in the market.

in the call option and short the underlying stock.

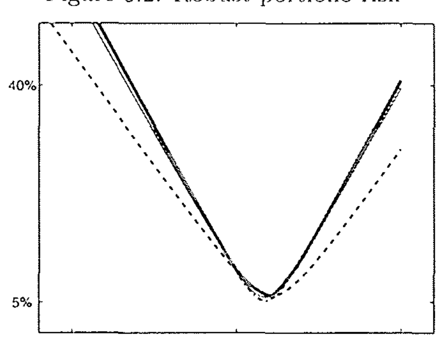
Figure 5.2 shows the risk of the portfolio as a function of w. Risk is defined as the standard deviation of the portfolio returns. The portfolio that maximizes the robust expected return is also a portfolio with very limited risk. In the example $\sigma = 10\%$, whereas the risk of the optimal portfolio is only around 8%. The qualifier "around" in assessing the risk is on purpose, since the risk is not uniquely defined. The portfolio variance depends on μ: If the option is out of the money r_k^d is fixed at -100%. Consequently

Figure 5.1: Robust expected portfolio return



Notes: The figure shows the expected portfolio return as a function of the investment w in the stock and an implicit investment of $1-w$ in the call option. The lines correspond to different beliefs about the expected stock return: $\mu = 2\%$ (dashed), $\mu = 5\%$ (thin line) and $\mu = 6\%$ (dash-dot). The thick curve corresponds to a robust evaluation of the expected portfolio return.

Figure 5.2: Robust portfolio risk



Notes: The figure shows the portfolio's standard deviation as a function of the investment w in the stock and an implicit investment of $1 - w$ in the call option. The lines correspond to different beliefs about the expected stock return: $\mu = 2\%$ (dashed), $\mu = 5\%$ (thin line) and $\mu = 6\%$ (dash-dot). The thick curve corresponds to a robust evaluation of the expected portfolio variance over all $\mu \in [\mu_L, \mu_U]$.

$r_k^d - \mu^d$ depends on μ and the portfolio variance is a quadratic function of μ.

In the remainder of the paper we show how to solve the robust portfolio problem in a much more general setting. We allow for multiple assets and options (puts and calls) with various strikes. The general model also allows for options on portfolios of stocks, like index or basket options. A second generalization concerns the uncertainty set $\mathcal{U}$. Instead of independent intervals on the expected returns of each asset, the uncertainty set will be characterized by an elliptical constraint to allow for correlation among the uncertainty in various assets. To handle such complicated uncertainty sets we develop a generalization of the robust transformation for linear constraints as discussed in section 1.3.1. We show that the general model can still be solved efficiently after reformulating the problem as an second order cone optimization problem (SOC). The solution technique is an extension of the interval partitioning in the univariate example. We provide an empirical illustration of the technique on a benchmark tracking problem.

We restrict to robust portfolio problems with piece-wise linear relations in the uncertain parameter. Consequently robust versions of variance constraints are not considered.

5.2 Portfolio returns without uncertainty

We consider the problem of allocating wealth over N different assets, possibly including the riskfree asset, and N^d options on these assets. Returns and expected returns on the risky assets are denoted by the N-vector r and μ respectively. The $(N \times N)$ covariance matrix of returns is denoted by Σ. The returns on the options depend on the returns of the risky assets. To make this functional form explicit one could use the N^d-vector $r^d(r)$ to denote option returns. For notational convenience we omit the argument r and use the N^d-vector r^d. A portfolio is a real $N + N^d$-vector $(w; w^d)$ with elements w_j which denote the amount of wealth allocated to the risky asset j and elements w_i^d which denote the wealth allocated to option i.

Our aim is to develop robust versions of linear portfolio constraints. An important type of linear constraint is a restriction on the expected portfolio performance

$$\mathsf{E}[r'w + r^{d\prime}w^d] \geq w_0. \tag{7}$$

Note that a study on such portfolio restrictions implicitly covers expected portfolio return maximization problems: we may avoid uncertain coefficients in the objective by reformulating the problem 'max $\mathsf{E}[r'w + r^{d\prime}w^d]$' as 'max$\{x : \ \mathsf{E}[r'w + r^{d\prime}w^d] \geq x\}$'.

In the remainder of this section, we introduce the notation to specify the option return vector r^d. We consider opportunity sets with one period options: we can buy the option, and if we do, keep it until expiration in the next time period. We adopt the usual notation in the financial literature that X denotes the exercise price, S denotes the price of the underlying asset when the option matures, and $S_0 > 0$ denotes the current price of the underlying asset. The return of the underlying asset is denoted $r := (S - S_0)/S_0$. Thus, X and S_0 are known quantities in non-negative real space R_+, whereas S and $1 + r$ are quantities in R_+ that are revealed at the next time epoch.

The payoff of a simple call option (the right to buy) with exercise price X is $\max\{0, S - X\}$. If the call option costs $c_0 > 0$, then its return is

$$r_c^d = \max\left\{0, \frac{S - X}{c_0}\right\} - 1.$$

Since $S = (1 + r)S_0$, we may rewrite r_c^d as a piece-wise linear function of r:

$$r_c^d = \max\{0, a_c r + b_c\} - 1 \text{ with } a_c := \frac{S_0}{c_0} \text{ and } b_c := \frac{S_0 - X}{c_0}.$$

with known coefficients $a_c > 0$ and b_c.

Similarly, the payoff of a simple put option (the right to sell) with exercise price X is $\max\{0, X - S\}$. If the put option costs $p_0 > 0$, then its return

$$r_p^d = \max\{0, a_p r + b_p\} - 1 \text{ with } a_p := -\frac{S_0}{p_0} \text{ and } b_p := \frac{X - S_0}{p_0}.$$

Consider N underlying assets (stocks and bonds) with unknown returns $r = (r_1, r_2, \ldots, r_N)'$ and N^d options with returns

$$r_i^d = \max\left\{0, b_i + \sum_{j=1}^{N} a_{ij} r_j\right\} - 1, \tag{8}$$

for some given b_i and a_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, N^d$. It is important to observe that (8) defines r^d as an explicit function of r_s. Call and put options on a single underlying asset k (simple options) correspond to the special case where $a_{ij} = 0$ for $j \neq k$ and $a_{ik} > 0$ or $a_{ik} < 0$ respectively. A basket option, which gives the right to buy (call) or sel (put) a basket of assets $N_b \subseteq N$ for a given price at maturity corresponds to a situation where $a_{ij} = 0$ for $j \notin N_b$ and $a_{ik} > 0\ \forall k \in N_b$ if it concerns a call or $a_{ik} < 0\ \forall k \in N_b$ if it concerns a put.

We say that the ith options is *in-the-money* if $1 + r_i^d > 0$ and *out-of-the-money* if $1 + r_i^d = 0$. The moneyness of the derivatives is determined by the realization of r through relation (8).

For any given realization r, the derivatives $\{1, 2, \ldots, N^d\}$ can be partitioned into the set $C \subseteq \{1, 2, \ldots, N^d\}$ of derivatives that are in-the-money, and the set $\tilde{C} := \{1, 2, \ldots, N^d\} \setminus C$ of derivatives that are out-of-the-money. Conversely, given a partition $(C, \tilde{C})$, $C \cup \tilde{C} = \{1, 2, \ldots, N^d\}$ and $C \cap \tilde{C} = \emptyset$, we let $P(C)$ be the set of returns that support the partition C, i.e.

$$P(C) := \{r :\ 1 + r_i^d > 0 \text{ for } i \in C,\ 1 + r_j^d = 0 \text{ for } j \in \tilde{C}\}. \tag{9}$$

As a matter of notation, we let $A \in R^{N^d \times N}$ denote the matrix with entries a_{ij} on the ith row and the jth column. Let A_C denote the $|C| \times N$ submatrix of A consisting of the rows $i \in C$, where $|C|$ denotes the cardinality of C. Similarly, we let $b_C \in R^{|C|}$ denote the subvector of b with entries b_i, $i \in C$. Thus, after a suitable row permutation we have

$$A = \begin{bmatrix} A_C \\ A_{\tilde{C}} \end{bmatrix}, \ b = \begin{bmatrix} b_C \\ b_{\tilde{C}} \end{bmatrix}.$$

By (8) and (9),

$$P(C) = \{r : \ b_C + A_C r > 0, \ b_{\tilde{C}} + A_{\tilde{C}} r \leq 0\}. \tag{10}$$

Observe that $P(C)$ is a polyhedral set. Furthermore, we have that

$$1 + r^d_C = b_C + A_C r \text{ for } r \in cl\ P(C) \tag{11}$$

and

$$1 + r^d_{\tilde{C}} = 0 \text{ for } r \in cl\ P(C). \tag{12}$$

Hence r^d is a linear function of the uncertain parameter r on $P(C)$, where C is an arbitrary partition of $\{1, 2, \ldots, N^d\}$. Strictly speaking, the set $\{1, 2, \ldots, N^d\}$ can be partitioned in 2^{N^d} ways, but most of these moneyness configurations have an empty set of supporting returns $P(C)$.

We consider the moneyness configurations $(C, \tilde{C})$ that partition the return space into non-empty sets $P(\cdot)$

$$\mathcal{F} := \{C : \ C \cup \tilde{C} = \{1, 2, \ldots, N^d\}, \ C \cap \tilde{C} = \emptyset, \ P(C) \notin \emptyset\}. \tag{13}$$

We are now able to specify the return on the option investment

$$r^{d\prime} w^d = (A_C r + b_C - \iota)' w^d_C \ \text{ if } r \in P(C) \tag{14}$$

with w^d_C the sub-vector of w^d corresponding to the options in the set C. Note that (14) is linear for each $C \in \mathcal{F}$.

For a special case with simple call and put options, the configurations are characterized by sorting derivatives on the same underlying according to the exercise price. Successive exercise prices X^l and X^u of the same underlying asset define a return interval $\left[\frac{X^l}{S_0}, \frac{X^u}{S_0}\right]$ for which the call options with $X \leq X_l$ and the put options with $X \geq X_u$ are in-the-money and included in C. We combine such intervals of alternative underlying assets to form a partition C of options such that $P(C)$ is non-empty,

$$P(C) = \left\{ r : \ \frac{X^l_{C,j}}{S_{0,j}} \leq 1 + r_j \leq \frac{X^u_{C,j}}{S_{0,j}}, \ \forall j = 1, \ldots, N : \right\},$$

with $X^{l}_{C,j}$ and $X^{u}_{C,j}$ denoting the successive exercise prices of options on underlying j, which specify the boundaries on the set $P(C)$ along dimension j. If we let N^d_j denote the number of derivatives on the underlying j, then there are at most $N^d_j + 1$ of these intervals for that particular asset and $\prod_{j=1}^{N}(N^d_j + 1)$ in total.

5.3 Linear portfolio restrictions with uncertainty

In practice the return distribution is not known precisely and investors rely on estimates which inevitably induce errors in the coefficients r (and consequently also $r^d(r)$) of the constraint. We consider an investor who is concerned about estimation uncertainty and aims at designing portfolios that are robust to this uncertainty. We contextualize the study by considering a robust investor who adopts a multivariate return model

$$r_k = \mu + \varepsilon_k \ \text{ with probability } \pi_k, \qquad k = 1, .., K. \tag{15}$$

and is concerned about the mean return vector μ.

Not all mean return vectors are plausible (according to the robust investor). The subset $\mathcal{U}$ of plausible return vectors of the N underlying assets is called the *uncertainty set*. In this section, we assume that $\mathcal{U}$ is the intersection of $\{\mu \in R^N : \ \mu \geq -1\}$ with an N-dimensional ellipsoid, i.e.

$$\mathcal{U} = \{\mu \in R^N : \ \mu \geq -1, \ \|A(\mu - \hat{\mu})\| \leq \theta\} \tag{16}$$

where A is a given $M \times N$ matrix (typically $M = N$), and θ is a given positive scalar constant. For example, $\hat{\mu}$ is the sample mean and A is the upper triangular part of the Choleski decomposition of Ω^{-1}, with $\Omega = \mathsf{E}\left((\hat{\mu} - \mu)(\hat{\mu} - \mu)'\right)$ such that $A'A = \Omega^{-1}$.

We first consider a stylized situation with a singular return distribution:

$$r = \mu + \epsilon, \qquad \epsilon = 0 \tag{17}$$

In this case a restriction (7) on the expected portfolio performance reduces to

$$r'w + r^{d'}w^d + w_0 \geq 0 \qquad \text{for all } r(= \mu) \in \mathcal{U}, \tag{18}$$

with the return on the option holdings defined by (14). The option return is a linear function of the unknown parameter r $(= \mu)$ on each set $P(C)$ but is non-linear on R^N, in particular it is piecewise linear with non-differentiable (adjoint) points given by $\{\mathrm{cl}(P(C_1)) \cap \mathrm{cl}(P(C_2)), \ C_1, C_2 \in \mathcal{F}\}$.

The uncertainty set $\mathcal{U}$ is not finite and in fact not countable. Hence (18) represents an infinite number of portfolio constraints on the design variables and is inadequate for computational reasons. Ideally, there exists a transformation of the infinite number of constraints (18) to a finite set of constraints which implies an equivalent feasible set for the design variables and which can be incorporated in an efficiently solvable optimization problem. We propose a method to transform (18) into finitely many second order cone restrictions

First, we partition the robust portfolio constraint (18) into finitely many robust linear constraints. As the constraint is linear in μ on the intervals $P(C)$, $C \in \mathcal{F}$ (see (14)), it is also linear on

$$\mathcal{U}(C) := \mathcal{U} \cap P(C)$$

for all $C \in \mathcal{F}_\mu = \mathcal{F} \cap \{C : \mathcal{U} \cap P(C) \neq \emptyset\}$. The moneyness configurations *partition* the uncertainty set $\mathcal{U}$ into at most $|\mathcal{F}_\mu|$ *ellipsoidal cuts*, i.e. $\cup_{C \in \mathcal{F}_\mu} P(C) \supseteq \mathcal{U}$. Hence (18) is equivalent to

$$w_0 + \mu' w + (A_C \mu + b_C - \iota)' w_C^d \geq 0 \qquad \text{for all } \mu \in \mathcal{U}(C),\ C \in \mathcal{F}_\mu. \tag{19}$$

and a solution to (19) satisfies (18) and vice versa.

Second, we design a transformation for a linear constraint in the uncertain parameter μ and an uncertainty set which is the intersection of a linear and ellipsoidal uncertainty set. On each ellipsoidal cut $\mathcal{U}(C)$, (19) is a linear (affine) function on μ, which we may rewrite as

$$f_{C,0}(w_0, w, w^d) + \sum_{j=1}^{N} f_{C,j}(w_0, w, w^d)' \mu \geq 0 \qquad \text{for all } \mu \in \mathcal{U}(C). \tag{20}$$

Since the coefficients of this function are different for each ellipsoidal cut $\mathcal{U}(C)$, we have added a subscript C. In particular, for $\mu \in \mathcal{U}(C)$

$$f_{C,0}(w_0, w, w^d) = w_0 + (b_C - \iota)' w_C^d \tag{21}$$

and

$$f_{C,j}(w_0, w, w^d) = w_j + \sum_{i \in C} w_i^d a_{ij} \text{ for } j = 1, 2, \ldots, n. \tag{22}$$

Moreover $f_C(w_0, w, w^d)$ is a N-vector with elements $f_{C,j}(w_0, w, w^d)$, $j = 1, .., N$, hence

$$f_{Cj}(w_0, w, w^d) = w + w_C^{d\,\prime} A_C$$

Define

$$\tilde{A} := \begin{bmatrix} A_C \\ -A_{\bar{C}} \\ I_N \end{bmatrix}, \ \tilde{b} := \begin{bmatrix} b_C - \iota \\ -b_{\bar{C}} + \iota \\ \iota \end{bmatrix},$$

and

$$P = \begin{bmatrix} 0^{\mathrm{T}} \\ A \end{bmatrix}, \ q = \begin{bmatrix} \theta \\ -A\hat{\mu}. \end{bmatrix},$$

By (10) and (16),

$$\mathcal{U}(C) = \left\{\mu \in R^N : \tilde{A}\mu + \tilde{b} \geq 0, \; P\mu + q \in \mathsf{SOC}\right\}.$$

Let

$$\mathcal{K}(C) := \{(f_0, f) : f_0 + f'\mu \geq 0 \text{ for all } \mu \in \mathcal{U}(C)\},$$

and note that (20) is equivalent to

$$\begin{bmatrix} f_{C,0}(w_0, w, w^d) \\ f_C(w_0, w, w^d) \end{bmatrix} \in \mathcal{K}(C). \tag{23}$$

Lemma 1 *The set of feasible solutions* (w_0, w, w^d) *to* (23) *has an equivalent formulation,*

$$\begin{array}{rcl} f_{C,0}(w_0, w, w^d) & \geq & q'u + \tilde{b}'v \\ f_C(w_0, w, w^d) & = & P'u + \tilde{A}'v \\ u & \in & SOC \\ v & \geq & 0. \end{array} \tag{24}$$

The lemma follows from theorems 6 and 7 presented in the appendix A.5.

We obtain a transformation for the robust portfolio constraint (18) if we use (24) for all $C \in \mathcal{F}_{\mathcal{U}}$. The transformation leads to an equivalent description of the feasible set which is now described by a finite number of second order cone constraints. For simple call and put options, we have at most $N_j^d + 1$ exercise intervals $[X^l, X^u]$ for asset j, with N_j^d denoting the options on underlying j. Hence $|\mathcal{F}_{\mathcal{U}}| \leq \prod_{j=1}^{N}(N_j^d + 1)$. For a fixed number of options, the transformation of (18) by lemma 1 is efficient.

Lemma 2 *If* A *is non-singular,* (20) *is equivalent to*

$$\begin{array}{c} w_0 + \hat{\mu}'w + (A_C'\hat{\mu} + b_C)'w_C^d - (\tilde{A}'\hat{\mu} + \tilde{b}')v \geq \theta \|(A')^{-1}(w + A_C^T w_C^d - \tilde{A}'v)\| \\ v \geq 0 \end{array} \tag{25}$$

Proof Use (21) and (22) to substitute for $f_{C,0}(w_0, w, w^d)$ and $f_C(w_0, w, w^d)$ in (24),

$$\begin{array}{rcl} w_0 + b_C' w_C^d & \geq & q'u + \tilde{b}'v \\ w + A_C' w_C^d & = & P'u + \tilde{A}'v \\ u & \in & \mathsf{SOC} \\ v & \geq & 0 \end{array} \tag{26}$$

As A is non-singular, we can express the equality in (26),

$$u_{2:N+1} = (A')^{-1}(w + A_C' w_C^d - \tilde{A}'v). \tag{27}$$

We substitute for $u_{2:N}$ in the first inequality which yields

$$w_0 + b_C' w_C^d \geq \theta u_1 - (A')^{-1}(w + A_C' w_C^d - \tilde{A}'v) + \tilde{b}'v$$

As $u \in SOC$, $u_1 \geq \|u_{2:N+1}\|$. We substitute

$$u_1 = \|u_{2:N+1}\| \tag{28}$$

to obtain the largest set of feasible solutions to the first inequality,

$$\begin{aligned} w_0 + \hat{\mu}'w + (A_C'\hat{\mu} + b_C)'w_C^d - (\tilde{A}'\hat{\mu} + \tilde{b}')v \\ -\theta \|A'^{-1}(w + A_C' w_C^d - \tilde{A}'v)\| \geq 0 \end{aligned} \tag{29}$$

For any feasible solution (w_0, w, w^d, v) to (29), with $v \geq 0$, we use (27) for $u_{2:N+1}$ and set $u_1 = \|u_{2:N+1}\|$. By construction $u \in SOC$ and (w_0, w, w^d, v, u) is a feasible solution to (26). Alternatively, if (w_0, w, w^d, v, u) is a feasible solution to (26), then $(w_0, w, w^d, \tilde{u})$ with $\tilde{u} = (\tilde{u}_1, u_{2:N+1})$, $\tilde{u}_1 \leq u_1$ such that $\tilde{u}_1 = \|u_{2:N+1}\|$, is also feasible to (26) and by (27)-(28) a feasible solution to (29). $\square$

If A follows from the Choleski decomposition of Ω^{-1}, then (25) reduces to

$$\begin{aligned} w_0 + \hat{\mu}'w + (A_C'\hat{\mu} + b_C)'w_C^d - (\tilde{A}'\hat{\mu} + \tilde{b}')v - \\ \theta\sqrt{(w + A_C^T w_C^d - \tilde{A}'v)'\Omega(w + A_C^T w_C^d - \tilde{A}'v)} &\geq 0. \\ v &\geq 0 \end{aligned} \tag{30}$$

Observe the similarity between (30) and the robust portfolio return (28) in chapter 2 when we exclude options, i.e. when A_C, b_C, $\tilde{A}$ and $\tilde{b}$ are zero. In fact, (30) is the generalization of the the robust portfolio constraint needed to handle an investment opportunity set with options.

So far we considered a return model with one possible realization $r = \mu$. Alternatively if the constraint is defined for a particular return $r_k = \mu + \varepsilon_k$. Constraint (19) changes to

$$w_0 + (\mu + \varepsilon_k)'w + (A_C(\mu + \varepsilon_k) + b_C - \iota)'w_C^d \geq 0 \qquad \text{for all } \mu \in \mathcal{U} \cap P_{\varepsilon_k}(C), \tag{31}$$

with

$$P_{\varepsilon_k}(C) = \{\mu : \ \mu + \varepsilon_k \in P(C)\}. \tag{32}$$

Moreover, we may consider the expected value of the portfolio under a discrete approximating return distribution (15). The robust version of a linear constraint on the expected value of the portfolio is

$$w_0 + \mu'w + \sum_{k=1}^{K} \pi_k r^d(\mu + \varepsilon_k)'w^d \geq 0 \text{ for all } \mu \in \mathcal{U}. \tag{33}$$

In this case we need a finer partitioning of $\mathcal{U}$. Let C_k denote a partition of the options associated with ε_k. The new, refined, partitioning considers the set C of all combinations of $C_1, C_2, .., C_K$ such that $\cap_{k=1}^{K} P_{\varepsilon_k}(C_k) \neq \emptyset$. For each element $\check{C} = \{C_1, C_2, .., C_K\}$ in C, the portfolio return is linear on the polyhedral set $\cap_{k=1}^{K} P_{\varepsilon_k}(C_k)$ and (33) is equivalent to imposing, for all elements in C,

$$w_0 + \mu' w + \sum_{k=1}^{K} \pi_k (A_{\check{C}}(\mu + \varepsilon_k) + b_{\check{C}} - \iota)' w_{\check{C}}^d \geq 0 \text{ for all } \mu \in \mathcal{U} \cap_{k=1}^{K} P_{\varepsilon_k}(C_k). \quad (34)$$

The constraints (31) and (34) remain linear and each uncertainty set remains an intersection of an ellipsoid and a polyhedral set. Hence we may apply theorems 6 and 7 to obtain a finite number of second order cones which may replace the infinite number of constraints.

For a portfolio constraint with N_j^d options on asset j and K scenarios, the maximum number of second order cone and linear constraints after the transformation is $\mathcal{O}(\prod_{j=1}^{N} K(N_j^d + 1))$. For N_j^d and K fixed, this is an 'efficient' transformation.

5.4 Benchmark tracking

We illustrate the presented methodology on a benchmark tracking problem: the decision maker is given a budget which may be invested in a portfolio consisting of assets and options on these assets to track a benchmark portfolio which does not belong to the investment set. To measure the success of tracking the benchmark we define the tracking error as the difference between the portfolio return and benchmark return.

The benchmark tracking problem is to design a portfolio which solves

$$\min_{w, w^d} \left\{ \mathsf{E}[(\tau(r, w, w^d))^2] : (w, w^d) \in \mathcal{W} \right\}. \quad (35)$$

where

- $w \in R^N$ and $w^d \in R^{N^d}$ are the portfolio investments in N assets and N^d options respectively,
- $r \in R^N$ is a vector denoting asset returns,
- $\mathsf{E}[\,]$ is an expectation operator conditional on the return distribution,
- $\tau(r, w, w^d) = r_0 - (r'w + r'w^d)$ is the difference between the benchmark return r_0 and the portfolio return,
- $\mathcal{W} = \{(w; w^d)' \subseteq R^{N+N^d} : (\sum_{i=1}^{N} w_i) + \sum_{j=1}^{N^d} w_j^d = 1\}$ models the budget for investment and may be supplemented with additional portfolio restrictions faced by the investor.

Suppose the return distribution is known. If the distribution is continuous, the expectation operator involves an N-dimensional integral which can be computed with reasonable accuracy using Monte Carlo methods,

$$\mathsf{E}[\tau^2(r, w, w^d)] \approx \sum_{k=1}^{K} \pi_k \tau^2(r_k, w, w^d) \tag{36}$$

with $(\pi_k, r_k) \in R_+ \times R^N$, $k = 1, .., K$ a sample from the return distribution such that $\sum \pi_k = 1$. If we use (36) to substitute for the objective, the problem reduces to,

$$\min_{(w;w^d)\in\mathcal{W}} \quad \sum_{k=1}^{K} \pi_k \tau^2(r_k, w, w^d) \tag{37}$$

This is the classical stochastic programming approach as described in Birge and Louveaux (1997).

If the mean return vector is unknown, then the tracking error $\tau^2(r_k, w, w^d)$ for a return r_k which is characterized by some deviation ε_k from the mean, is also unknown. To make the functional relationship explicit, we will now write $r(\mu)$. We consider a robust investor who considers, for each realization k, the worst squared tracking error that is possible for $\mu \in \mathcal{U}$, i.e. the robust investor solves[3]

$$\begin{array}{ll} \min_{(w;w^d)\in\mathcal{W}} & \sum_{k=1}^{K} \pi_k \tau_k^2 \\ & \tau_k^2 \geq \tau^2(r_k(\mu), w, w^d) \qquad \forall \mu \in \mathcal{U}, \qquad \forall k = 1, .., K. \end{array} \tag{38}$$

To obtain an optimization problem that is suitable for applying the previously developed methodology, we use an alternative description of the feasible set of (38):

$$\begin{array}{rcllll} \tau_k & \geq & \tau(r_k(\mu), w, w^d) & \forall \mu \in \mathcal{U}, & \forall k = 1, .., K \\ \tau_k & \geq & -\tau(r_k(\mu), w, w^d) & \forall \mu \in \mathcal{U}, & \forall k = 1, .., K. \end{array} \tag{39}$$

After reformulation, the constraints are

$$\begin{array}{rcll} \tilde{r}(\mu)'\tilde{w} + r^{d\prime}w^d + \tau_k & \geq & 0 & \forall \mu \in \mathcal{U} \\ \tilde{r}(\mu)'\tilde{w} + r^{d\prime}w^d - \tau_k & \leq & 0 & \forall \mu \in \mathcal{U} \end{array} \tag{40}$$

with $\tilde{r}(\mu) = (r_0; r(\mu))$, $\tilde{w} = (-1; w)$. This type of constraint is covered by (31) for which a transformation exists.

If we use return model (15) then P, as defined in (32) partitions the uncertainty set $\mathcal{U}$. On each partition (40) presents an infinite set of linear restrictions and we may apply theorems 6 and 7 to obtain a finite number of restrictions that are equivalent to the constraints in (38).

[3]Note that this investor is more conservative than a robust investor who solves $\min_{w,w^d} \max_{\mu\in\mathcal{U}} \sum_{l=1}^{K} \pi_l \tau^2(r_l(\mu), w, w^d)$.

Alternatively one could assume that returns are lognormally distributed, i.e.

$$\ln r = \ln \mu + \varepsilon \tag{41}$$

such that $r = e^{\ln \mu + \varepsilon} = \mu.e$ with $e = e^{\varepsilon}$ and the portfolio return is

$$(\mu.e)'w + r^d((\mu.e))'w^d \tag{42}$$

and is piecewise linear in μ. Moreover (42) is linear in μ on the polyhedral set

$$P_{\varepsilon_k}(C) = \{\mu : \ \mu.\varepsilon_k \in P(C)\}. \tag{43}$$

Hence the assumptions, needed to apply the results of section 5.3 hold.

5.5 Empirical study

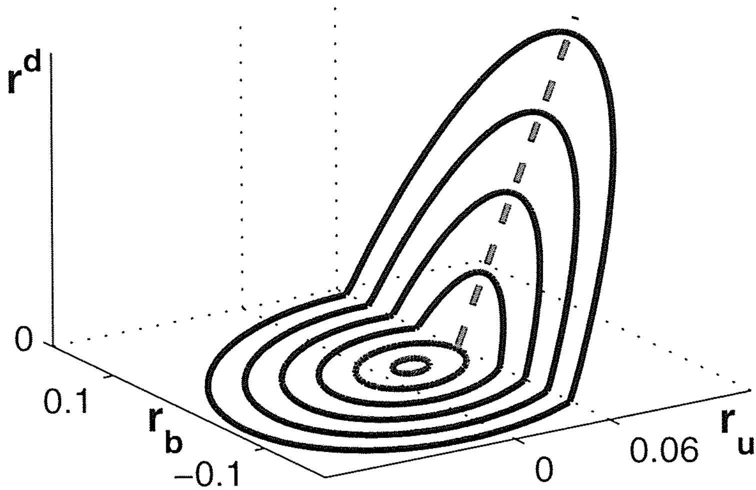
We illustrate the robust approach to the benchmark tracking problem with options and compare it to a naive approach. The Dow-Jones index serves as benchmark and the investment set consists of the Eurex stoxx 50 index and all options on this index.

To obtain consistent data, in particular consistent option prices, we use the historical data on monthly Dow-Jones returns (r_b), monthly Eurex stoxx 50 returns (r_s) and option prices for each month over the period March 1997 to March 2002. Table 5.1 reports the return characteristics. With a bootstrap experiment on this dataset we analyze the robust approach. We draw M bootstrap samples from the dataset. Each bootstrap sample consists of T historical return observations of the benchmark (Dow-Jones index) and the underlying asset (Eurex index). The date corresponding to the last observation in the bootstrap $m = 1, .., M$ is denoted t_m. Also a set of options prices which mature at $t_m + 1$ is added to the sample information. More specifically, we consider 20 options with exercise prices closest to the to the estimated stock price for the next period.

For each of the M bootstrap samples we compute the optimal naive and robust portfolio allocation as follows:

- We estimate a log-normal return model (41) on the bootstrap sample.
- We approximate the two-variate return model by a stratified sample $\{\pi_k, \varepsilon_k\}_{k=1}^{K}$ with $K = 250$ scenarios such that $r_k = \mu + \varepsilon_k$, $k = 1, .., K$ cover all exercise intervals when μ is set to its estimate.
- The uncertainty set is based on the sample means $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$ of the bootstrap sample. We set $\Omega = \hat{\Sigma}/T$ and $\theta = \chi^2_{inv}(N, p)$ in the definition of the set $\mathcal{U}$. For a naive approach without considering uncertainty, we set $\theta = 0$. For the robust investor we use $N = 2$ and $p = 0.95$ and obtain $\theta = 2.4$.

Figure 5.3: Uncertainty set for two assets and an option



Notes The figure shows uncertainty sets for the returns that follow from a given deviation ε from the uncertain mean μ. The returns r_b, r_u and r^d correspond to a benchmark (Dow-Jones index), an underlying asset (Eurex stoxx 50 index) and a call option on the underlying asset with exercise price $1.06S_0$ respectively. The solid black lines depict the boundaries of uncertainty sets with alternative levels of robustness θ: $\theta = 0.2$ for the smallest ellipsoid and $\theta = 2.4$ for the largest figure. The dotted line corresponds to the call's return, $r^d = ar_s + b$, if the option is in-the-money.

- We formulate the robust tracking model (38) with

$$\mathcal{W} = \{(w; w^d) : \ w = (w_b; w_s) \in (-1) \times R, \ w^d \in R^{N^d} \text{ and } w_s + \iota' w^d = 1\}$$

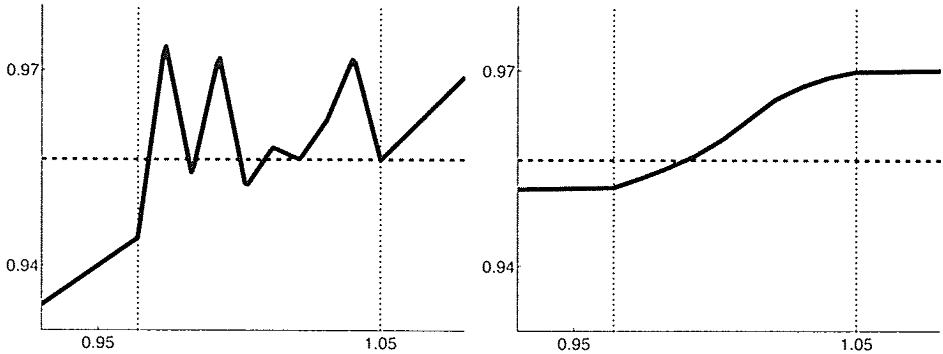
- We use (39)-(40) and lemma 2 to transform the problem to a second order cone problem[4]. We solve the problem numerically with SeDuMi 1.05 developed by Sturm (1999).

- We store the naive portfolio $(w_n; w_n^d)$ and robust portfolio $(w_r; w_r^d)$. We also store the ex-ante expected tracking errors $\hat{\tau}$ given by the objective value of the naive and robust solutions.

- We compute the ex-post tracking error τ based on the returns r_{b,t_M+1}. For example $\tau(w_r; w_r^d) = r_{b,t_M+1} - (w_r r_{s,t_M+1} + r^d(r_{s,t_M+1}) w_r^d)$.

[4]The problem has at most $2K(N^d + 1)$ second order cone constraints of dimension $(N + 1)$, a second order cone of dimension $K + 1$ to compute the objective value and maximally $1 + 2K(N^d + 1)N$ linear constraints. Numerical tests show that the dense formulation (25) of the constraint leads to less numerically problems and solves faster than the larger sparse formulation (26).

Figure 5.4: Portfolio returns



Notes The figure shows the portfolio revenue (one plus the return) as a function of the underlying asset's return for a typical naive portfolio allocation (left-panel) and a typical robust portfolio allocation (right-panel). The estimated mean return vector for this instance is $(1 + r_b, 1 + r_s) = (0.956, 0.975)$. The dashed horizontal line presents the mean expected benchmark return and serves as reference point as the objective is to minimize the tracking error. The vertical dotted lines present the smallest and largest exercise prices associated with the options in the investments set.

The following results should be interpreted with caution as our computational study is based on limited dataset with only 5 years of monthly observations. We observe some structural differences between the approaches and we describe them below.

Table 5.2 reports the results of the bootstrap. The naive investment in the underlying asset varies considerably over the bootstrap samples and is small compared to the investment (short or long) in the options. On average 0.04 is invested in the underlying asset and the average cross-sectional standard deviation is 3.40. Figures 5.4 and 5.5 show a typical example of portfolio returns. The naive portfolio aims to stay near the mean benchmark return conditional on the underlying's return: near the average underlying return and between the smallest and largest exercise limits of the options, the average benchmark return is copied by long and short positions in options that lead to the 'sawtooth' pattern with spikes at the underlying asset's returns that correspond to the exercise prices of options. When the value of the underlying asset is larger than its average, the portfolio also produces a larger return to mimic the benchmark return that

Table 5.1: Monthly return statistics

	μ	σ
Dow Jones return r_b	0.31%	7.47 %
EUREX return r_u	2.21%	6.06 %

Notes The table reports the means μ and standard deviations σ of the returns estimated on the period March 1997 to March 2002. The correlation between the assets is 25%.

Table 5.2: Results

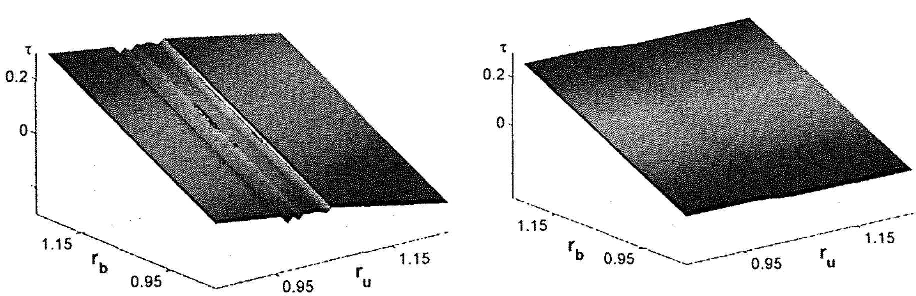
	Naive Approach $(w_n; w_n^d)$	Robust Approach $(w_r; w_r^d)$	Eurex only $w^d = 0$
ex-ante			
$E(\|\tau\|)$	7.3	11.6	7.6
	(0.3)	(5.4)	(2.0)
$E(\|\tau\|)$		8.7	
		(2.0)	
ex-post			
$E(\| \tau\|\|)$	12.0	9.1	7.6
	(13.5)	(6.3)	(5.6)
$\tau_{5\%}$	40.4	23.5	21.9
$P(\tau \leq \hat{\tau})$	43.9	64.8	75.0
portfolio			
investment underlying (w_s)	-0.38	1.02	1.00
	(4.30)	(0.11)	(0.00)
cross-sectional standard deviation	2.81	0.46	0.00
	(3.86)	(0.48)	(0.00)

Notes: The results are based on a bootstrap of size $M = 1000$, $T = 6$ observations and $K = 250$ scenarios. The columns correspond to different portfolios: the naive portfolio $(w_n; w_n^d)$ $(\theta = 0)$, the robust portfolio $(w_r; w_r^d)$ with $\theta = 2.4$ and w_d with an investment of 1 in the Eurex stoxx 50 index. All number are percentages per month.

is believed to be positively correlated to the underlying asset's return.

The ex-ante portfolio tracking errors of the naive portfolio are naturally smaller than the ex-ante tracking errors of a fixed strategy that invests only in the Eurex stoxx 50 index. However, ex-post it has considerably larger tracking error.

Figure 5.5: Tracking error



Notes The figure shows the ex-ante tracking error of the naive portfolio (left-panel) and robust portfolio (right-panel). The r_b and r_u axes denote the benchmark and underlying asset's revenue (one plus the return) and span approximately 99% of the probability mass. Otherwise the specifications of figure 5.4 apply.

The robust portfolio remains close to the underlying asset and has small investment in the risky options. More moderate options investments lead to a 'smooth' portfolio return in figure 5.4. Unlike the naive approach which tries to copy the benchmark return at every point, the robust approach aims for a portfolio which has smallest tracking error for any (r_b, r_s) return combination in the neighborhood, modelled by the uncertainty set, of each scenario. The slope of the robust portfolio return is smaller than its naive counterpart as the uncertainty in the mean also induces uncertainty in the correlation of the benchmark and underlying asset's return.

The robust portfolio has largest ex-ante tracking error, but has ex-post behavior that is superior to its naive competitor. The 5th percentile and even the expected tracking error are also smaller for the robust approach. The reliability, $P(\tau \leq \hat{\tau})$, of the robust portfolio is better than for the naive portfolio, but still quite low. Presumably, distributional properties of this small dataset are different than assumed and therefore the considered uncertainty set is not appropriate. On the drawback side, the robust portfolio performs inferior to a fixed investment in the Eurex index.

5.6 Extensions

For the numerical example of section 5.1, we argued that a transformation of a robust version of the variance constraint is not trivial. A study on efficient transformations would be an interesting venue for further research. Provided that we can robustly handle variance restrictions we could consider robust mean-variance portfolio choice with options. For an empirical analysis one could incorporate the bid-ask spread of the option prices to preclude arbitrage opportunities.

Another extension to this study would be to consider multi-period portfolio choice with the possibility to buy and sell options at any time. This requires the pricing of options at intermediate time periods. Models for option pricing such as the Black and Scholes (1973) option pricing formula lack precision to be adopted blindly and moreover rely on uncertain parameters such as the future stock price and volatility. Nevertheless option pricing models agree about the dependencies of option prices on different parameters. An approach for robust multi-period portfolio choice that acknowledges the relations between option prices and parameters but also the uncertainty in the parameters could consider an approximation of the option price dynamics and attribute imprecision to uncertainty in the parameters.

For example a first order approximation to a call option's price is

$$\tilde{c}(S, \sigma) = c(\hat{S}, \hat{\sigma}) + \delta(S - \hat{S}) + \mathcal{V}(\sigma - \hat{\sigma}) \tag{44}$$

with σ the underlying asset's standard deviation, $c(\hat{S}, \hat{\sigma})$ the Black and Scholes (1973) option price and the 'greeks' δ, Γ and $\mathcal{V}$ measure the option's price sensitivity to stock price, its curvature and stock volatility respectively. This approximation copies (locally) the dynamics of the Black and Scholes (1973) model. A robust approach results if we consider the worst option price (44) over all plausible expected stock returns and standard deviations.

A.5 Duality to achieve standard-form expressions

This appendix derives the duality result that we use in section 5.2.

Given a nonempty set $D \subseteq R^n$, its homogenized cone in R^{n+1} is defined as

$$\mathcal{H}(D) := cl\ \{(s, y) | s > 0,\ y/s \in D\}.$$

A set $\mathcal{K} \subseteq R^n$ is a cone if and only if $\mathcal{K} \neq \emptyset$ and

$$x \in \mathcal{K},\ t \geq 0 \implies tx \in \mathcal{K}.$$

If in addition,

$$x, y \in \mathcal{K} \implies x + y \in \mathcal{K}$$

then $\mathcal{K}$ is a convex cone. It is easily verified that $\mathcal{H}(D)$ is a cone; if D is convex then $\mathcal{H}(D)$ is a convex cone. The dual of a cone $\mathcal{K} \subseteq R^n$ is defined as

$$\mathcal{K}^* := \{s \in R^n | x^{\mathrm{T}} s \geq 0 \text{ for all } x \in \mathcal{K}\}.$$

A dual cone is always closed and convex. If $\mathcal{K}$ is convex, then the bi-polar relation holds:

$$(\mathcal{K}^*)^* = cl\ \mathcal{K}. \tag{45}$$

A second order cone (or Lorentz cone) is defined as

$$SOC = \left\{ x \in R^n :\ x_1 \geq \sqrt{x_2^2 + x_3^2 + \cdots x_n^2} \right\},$$

where n is the dimension of the second order cone. The interior of the second order cone is denoted int(SOC), i.e.

$$\text{int}(SOC) = \left\{ x \in R^n :\ x_1 > \sqrt{x_2^2 + x_3^2 + \cdots x_n^2} \right\}.$$

In the proof of Theorem 6 below, we need the following technical lemmas.

Lemma 3 *Let $D \neq \emptyset$. It holds that*

$$\mathcal{H}(D)^* = \{(f_0, f) \mid f_0 + f^{\mathrm{T}} r \geq 0 \text{ for all } r \in D\}.$$

The above lemma is a special case of Corollary 1 in (Sturm and Zhang 2003).

Lemma 4 *Let $\mathcal{K} \subseteq R^n$ be a cone and B an $m \times n$ matrix. Then*

$$\{x \mid Bx \in \mathcal{K}^*\} = \{B^{\mathrm{T}} y \mid y \in \mathcal{K}\}^*.$$

For a proof, see relation (17) in (Sturm and Zhang 2003). An important special case is that for two cones $\mathcal{K}_1$ and $\mathcal{K}_2$ one has

$$\mathcal{K}_1^* \cap \mathcal{K}_2^* = (\mathcal{K}_1 + \mathcal{K}_2)^*, \tag{46}$$

as obtained by setting $B := \begin{bmatrix} I, & I \end{bmatrix}^{\mathrm{T}}$ and $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$.

Lemma 5 *Let*

$$D = \{r \mid Pr + q \in SOC,\ \tilde{A}r + \tilde{b} \geq 0\}.$$

If $D \neq \emptyset$ then

$$\mathcal{H}(D) = \{(s, y) | Py + sq \in SOC,\ s \geq 0,\ \tilde{A}y + s\tilde{b} \geq 0\}.$$

Proof From the definition of $\mathcal{H}(D)$, it is clear that if $(s, y) \in \mathcal{H}(D)$ then

$$Py + sq \in \mathrm{SOC},\ s \geq 0,\ \tilde{A}y + s\tilde{b} \geq 0 \tag{47}$$

Conversely, suppose that (s, y) satisfies (47). If $s > 0$ then $y/s \in D$ and hence $(s, y) \in \mathcal{H}(D)$. Suppose now that $s = 0$. Since $D \neq \emptyset$, there exists $r \in D$. Let $\sigma > 0$ be arbitrary. We have from the definition of D and (47) that

$$P(r + \frac{1}{\sigma}y) + q \in \mathrm{SOC},\ \tilde{A}(r + \frac{1}{\sigma}y) + \tilde{b} \geq 0.$$

Hence $(\sigma r + y)/\sigma \in D$ and $(\sigma, \sigma r + y) \in \mathcal{H}(D)$ for all $\sigma > 0$. Letting $\sigma \downarrow 0$ it follows that $(0, y) \in \mathcal{H}(D)$. $\square$

Theorem 6 *Let*

$$D = \{r \mid Pr + q \in SOC,\ \tilde{A}r + \tilde{b} \geq 0\}$$

and consider the cone of linear functions that are nonnegative on D, i.e.

$$\mathcal{K} = \{(f_0, f) \mid f_0 + f^{\mathrm{T}} r \geq 0 \text{ for all } r \in D\}.$$

If $D \neq \emptyset$ then

$$\mathcal{K} = \mathrm{cl}\left\{ \begin{bmatrix} q^{\mathrm{T}} u + \tilde{b}^{\mathrm{T}} v + v_0 \\ P^{\mathrm{T}} u + \tilde{A}^{\mathrm{T}} v \end{bmatrix} \middle| u \in SOC,\ v \geq 0,\ v_0 \geq 0 \right\}.$$

Proof We have from Lemma 3 that

$$\mathcal{K} = \mathcal{H}(D)^*.$$

Applying Lemmas 5 and 4 respectively, we have

$$\begin{aligned}\mathcal{H}(D) &= \{(s,y) | Py + sq \in SOC\} \cap \left\{(s,y) \middle| \begin{bmatrix} 1 & 0^{\mathrm{T}} \\ \tilde{b} & \tilde{A} \end{bmatrix} \begin{bmatrix} s \\ y \end{bmatrix} \geq 0 \right\} \\ &= \left\{ \begin{bmatrix} q^{\mathrm{T}}u \\ P^{\mathrm{T}}u \end{bmatrix} \middle| u \in SOC \right\}^* \cap \left\{ \begin{bmatrix} v_0 + \tilde{b}^{\mathrm{T}}v \\ \tilde{A}^{\mathrm{T}}v \end{bmatrix} \middle| v_0 \geq 0,\, v \geq 0 \right\}^* .\end{aligned}$$

(It is well known that SOC and R^n_+ are self-dual cones.) Further using (46) and (45), we have

$$\mathcal{H}(D)^* = \mathrm{cl} \left\{ \begin{bmatrix} q^{\mathrm{T}}u + \tilde{b}^{\mathrm{T}}v + v_0 \\ P^{\mathrm{T}}u + \tilde{A}^{\mathrm{T}}v \end{bmatrix} \middle| u \in SOC,\, v \geq 0,\, v_0 \geq 0 \right\}.$$

□

The following theorem states that the closure operator in the above characterization of $\mathcal{K}$ is redundant if a Slater condition holds.

Theorem 7 *Let*

$$D^o := \{r \mid Pr + q \in \mathrm{int}(SOC),\ \tilde{A}r + \tilde{b} > 0\}$$

and let $\mathcal{K}$ be defined as in Theorem 6. If $D^o \neq \emptyset$ then

$$\mathcal{K} = \left\{ \begin{bmatrix} q^{\mathrm{T}}u + \tilde{b}^{\mathrm{T}}v + v_0 \\ P^{\mathrm{T}}u + \tilde{A}^{\mathrm{T}}v \end{bmatrix} \middle| u \in SOC,\, v \geq 0,\, v_0 \geq 0 \right\}.$$

Proof Let

$$\Gamma = \left\{ \begin{bmatrix} q^{\mathrm{T}}u + \tilde{b}^{\mathrm{T}}v + v_0 \\ P^{\mathrm{T}}u + \tilde{A}^{\mathrm{T}}v \end{bmatrix} \middle| u \in SOC,\, v \geq 0,\, v_0 \geq 0 \right\}.$$

We know from Theorem 6 that $\mathcal{K} = cl\ \Gamma$. It remains to show that Γ is closed, i.e. $cl\ \Gamma = \Gamma$. Let $(t,x) \in \mathcal{K} = cl\ \Gamma$, and let $(u^{(k)}, v^{(k)}, v_0^{(k)})$, $k = 1, 2, \ldots$ be a sequence such that

$$u^{(k)} \in SOC,\ v^{(k)} \geq 0,\ v_0^{(k)} \geq 0 \text{ for all } k = 1, 2, \ldots$$

and

$$\begin{bmatrix} t \\ x \end{bmatrix} = \lim_{k \to \infty} \begin{bmatrix} q^{\mathrm{T}}u^{(k)} + \tilde{b}^{\mathrm{T}}v^{(k)} + v_0^{(k)} \\ P^{\mathrm{T}}u^{(k)} + \tilde{A}^{\mathrm{T}}v^{(k)} \end{bmatrix}.$$

By definition of Γ, such a sequence must exist, because $(t,x) \in cl\ \Gamma$. Let $r \in D^o$. We have

$$\begin{aligned} t + r^{\mathrm{T}}x &= \lim_{k\to\infty} q^{\mathrm{T}}u^{(k)} + \tilde{b}^{\mathrm{T}}v^{(k)} + v_0^{(k)} + r^{\mathrm{T}}(P^{\mathrm{T}}u^{(k)} + \tilde{A}^{\mathrm{T}}v^{(k)}) \\ &= \lim_{k\to\infty} (Pr+q)^{\mathrm{T}}u^{(k)} + (\tilde{A}r + \tilde{b})^{\mathrm{T}}v^{(k)} + v_0^{(k)} \\ &\geq \lim_{k\to\infty} (Pr+q)^{\mathrm{T}}u^{(k)} + (\tilde{A}r + \tilde{b})^{\mathrm{T}}v^{(k)}. \end{aligned} \tag{48}$$

Since $Pr + q \in \mathrm{int}(SOC)$ and $\tilde{A}r + \tilde{b} > 0$, we have

$$\begin{cases} (Pr+q)^{\mathrm{T}}u > 0 \text{ for all } u \in SOC\backslash\{0\} \\ (\tilde{A}r + \tilde{b})^{\mathrm{T}}v > 0 \text{ for all } v \in R_+^n\backslash\{0\}. \end{cases} \tag{49}$$

We claim that the sequence $\{u^{(k)}\}$ is bounded. Indeed, suppose to the contrary that $\limsup_{k\to\infty} \|u^{(k)}\| = \infty$. From (49), it follows that

$$\liminf_{k\to\infty} \frac{(Pr+q)^{\mathrm{T}}u^{(k)}}{\|u^{(k)}\|} > 0,$$

and hence, using also (48), we arrive at the impossible inequality

$$t + r^T x \geq \limsup_{k\to\infty} (Pr+q)^{\mathrm{T}}u^{(k)} = \infty.$$

Similarly, we can show by contradiction from (48) and (49) that $v^{(k)}$ must be bounded. Hence this sequence $\{u^{(k)}, v^{(k)}, v_0^{(k)}\}$ has a cluster point (u, v, v_0), $u \in SOC$, $v \geq 0$, $v_0 \geq 0$, and

$$\begin{bmatrix} t \\ x \end{bmatrix} = \begin{bmatrix} q^{\mathrm{T}}u + \tilde{b}^{\mathrm{T}}v + v_0 \\ P^{\mathrm{T}}u + \tilde{A}^{\mathrm{T}}v \end{bmatrix} \in \Gamma.$$

This concludes the proof. □

Chapter 6

Conclusion

The motivation for this thesis is the poor (mean-variance) portfolio performance of naive portfolio choice. A naive investor disregards fundamental uncertainty in the return model: The investor chooses a return model, i.e. selects a model specification and estimates the associated parameters. Based on this return model, she chooses the optimal portfolio. Possible return model misspecification and estimation errors in the parameters lead to uncertainty in the return model. The naive investor does not consider the uncertainty. But if the model is misspecified or estimation errors arise, the naive portfolio is actually suboptimal. The problem is that small changes in the return model, which is the input for the portfolio optimization, lead to a fundamentally different optimal portfolio choice. Different portfolios have different actual performance. Preceding studies and chapters 2 and 3 of this thesis show that this effect of uncertainty leads to a dramatic actual portfolio performance for naive portfolio choice.

More reliable models to describe return behavior and a larger number of observations could reduce uncertainty and improve the performance. A lot of effort has been put into developing reliable models to describe return behavior. Yet, an investor is still pressed for a procedure which translates all available information and its accompanying uncertainty into a reliable decision. One could consider a Bayesian approach and maximize the weighted performance over alternative return models. This would reduce the sensitivity of the portfolio on any specific return model. Nevertheless, the specific weights for the alternative models need to be chosen and will influence the portfolio choice.

In this thesis we consider a robust approach to deal with uncertainty: The robust investor optimizes for a portfolio which has the best worst case performance where the worst case is drawn from a limited set of return models. This set reflects the investor's view on uncertainty. The worst case approach inspires the investor's confidence that the computed performance will actually be obtained. However the investor does not intend to be conservative. Therefore the investor only includes return models which are, conditional on the data, plausible and among which she does not know how to choose. We consider among others statistical (Bayesian) methods to quantify the set of plausible return models.

We stress that robustness is not incompatible with expected performance maximization in the presence of uncertainty: When the return model is given with certainty, the robust and naive performance maximizing solution coincide. On the other hand, uncertainty about the return model may reduce the information available to a set of plausible models. As the models are all plausible and, in a statistical sense, indistinguishable there is no ground for selecting any particular model as the basis for portfolio choice. Instead a robust investor pursues the maximal expected performance that can be guaranteed across all the plausible models. This does not detract the objective from expected performance optimization but is an alternative to choosing just any return model as a basis for the decision. Without considering uncertainty the expected performance is typically overestimated and overconfident portfolio choice results (chapters 2 and 3).

Robust counterparts to optimization problems are computationally more difficult to solve than the original, non-robust version of the problem. When the uncertainty set contains a finite number of elements, the robust optimization problem maintains the structure of the original problem but the dimensions of the problem are scaled up. A non-finite uncertainty set is more problematic. In this case, the computational tractability crucially depends on the possibility to transform the robust optimization problem to an equivalent finite dimensional problem. For some, empirical relevant, special cases of uncertainty we were able to derive analytical results to single period portfolio choice problems, in the other cases we show that the robust portfolio allocation may be computed efficiently by numerical optimization (chapters 2 and 5). We studied robust multi-period portfolio choice in chapter 4. There the combination of uncertainty, return model and objective exclude an efficient transformation of the robust problem. We solve the problem by numerical methods but conclude that computation time is considerable. It would be interesting to study approximations of this problem which are easier to solve, for example a finite approximation to the uncertainty set. In this context, but also motivated by impasses in this dissertation which we resolved by using an approximation, it would be interesting to find a transformation of uncertain portfolio constraints which are bi-linear in the uncertain parameters.

From the analytical results, we find that three considerations dominate robust one-period mean-variance portfolio choice. 1) The robust solution uses a preliminary test to assess whether the risky market offers reliable investment opportunities. 2) When the market stands the preliminary test, the robust investor enters the risky market but will invest less in risky assets than a naive investor. 3) The robust investor changes the composition of the risky asset allocation compared to the naive asset allocation when the measure for uncertainty deviates from the measure of risk. For robust portfolio choice under a benchmark constraint we find that the robust solution tends to zero risky investment if the market does not offer reliable investment opportunities (figure 2.2).

We find that also a robust multi-period investor holds less risky assets than a naive or Bayesian multi-period investor and refrains from risky investment if uncertainty is too large. The computation of the robust solution also indicates which aspects of the return model are crucial for portfolio choice. Depending on the initial state of the economy and the investment horizon, crucial parameters are the persistence of the initial value of the predictor variable, the long term average of the predictor variable and the extent of predictability in asset returns.

The empirical results show that - in the presence of uncertainty - a robust approach has considerably better actual portfolio performance than a naive approach. Chapters 2, 3 and 5 show that a significant gain in robustness can be attained at the expense, if any, of only a small decrease in expected performance. More than that, we observe that a robust approach to estimation uncertainty leads to better expected performance than a naive approach. Especially the implicit preliminary test of the investment opportunity set proves to be an effective guard against error-maximizing portfolios.

An estimation robust portfolio based on an unstructured return model has a performance which is equivalent to the performance of approaches that reduce the effect of estimation uncertainty by imposing structure on the return model, e.g. the CAPM and the Fama & French asset pricing model (chapter 3). The incidental advantage of the robust approach is that its ex-ante expected performance is a robust estimator of its ex-post performance.

The model robust solution to mean-variance portfolio choice is the mean-variance portfolio choice associated with an endogenously determined combination of the alternative plausible models. The results show (table 3.2) that model robustness may even improve upon the performance of the best performing alternative model. We considered several prominent alternative models to describe uncertainty; it would be interesting to consider the effect of model robustness if this set is extended, e.g. by applying a Bayesian model selection approach.

The results for one-period portfolio choice suggest that the robust investor is a formidable competitor to the naive investor for maximizing actual portfolio performance in the presence of uncertainty. The virtues of robust multi-period portfolio choice remain to be explored by future research.

Bibliography

Aït-Sahalia, Y. and Brandt, M. 2001. Variable selection for portfolio choice. *Journal of Finance* **56**(4), pp. 1297–1351.

Alizadeh, F., Haeberly, J., Nayakkankuppam, M. and Overton, M.: 1997. *SDPPACK User's Guide, Version 0.8 Beta.* NYU.

Anderson, E.: 2000. *The MOSEK base system and application program, interface version 1.3.* EKA Consulting ApS. Denmark.

Anderson, E., Hansen, L. and Sargent, T.: 1999. Robustness, detection, and the price of risk. Working paper, Stanford University.

Anderson, T.: 2003. *An introduction to multivariate statistical analysis.* 3 edn. Wiley. New York.

Avramov, D. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* **64**(3), pp. 423–458.

Barberis, N. 2000. Investing for the long run when returns are predictable. *The Journal of Finance* **55**(1), pp. 225–264.

Ben-Tal, A. and Nemirovski, A. 1997. Robust truss topology design via semidefinite programming. *SIAM Journal Optimization* **7**, pp. 991–1016.

Ben-Tal, A. and Nemirovski, A. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**, pp. 769–805.

Ben-Tal, A. and Nemirovski, A.: 2001. *Lectures on modern convex optimization: analysis methods and engineering applications.* SIAM-MPS.

Ben-Tal, A., Margalit, T. and Nemirovski, A.: 2000. Robust modeling of multi-stage portfolio problems. *in* H. Frenk, K. Roos, T. Terlaky and S. Zhang (eds), *High Performance Optimization.* Kluwer Academic Publishers. chapter 12, pp. 303–328.

Ben-Tal, A., Nemirovski, A. and Roos, C. 2001. Robust solutions of uncertain quadratic and conic-quadratic problems. *Working Paper.*

Birge, J. and Louveaux, F.: 1997. *Introduction to stochastic programming.* Springer.

Black, F. and Litterman, R.: 1990. Asset allocation: Combining investor views with market equilibrium. *Technical report.* Goldman, Sachs & Co., Fixed Income Research.

Black, F. and Litterman, R. 1992. Global portfolio optimization. *Financial Analysts Journal* **48**(9), pp. 28–43.

Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, pp. 637-654.

Brennan, M. 1998. The role of learning in dynamic portfolio decisions. *European Financial Review* **1**, pp. 295–306.

Campbell, J. and Viceira, L.: 2002. *Strategic asset allocation: portfolio choice for long term investors.* Oxford University Press.

Campbell, J., Lo, A. and Mackinley, A.: 1997. *The econometric of financial markets.* Princeton Unversity Press.

Cavadini, F., Sbuelz, A. and Trojani, F.: 2001. A simplified way of incorporating model risk, estimation risk and robustness in mean variance portfolio management. Working Paper CentER, Tilburg University, The Netherlands.

Chamberlain, G. 2000. Econometric applications of maxmin expected utility. *Journal of Applied Econometrics* **15**, pp. 625–644.

Chen, Z. and Epstein, L. 2002. Ambiguity, risk and asset returns in continuous time. *Econometrica* **70**, pp. 1403–1443.

Chopra, V. and Ziemba, W. 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management* **19**(2), pp. 6–12.

Costa, O. and Paiva, A. 2002. Robust portfolio selection using linear-matrix inequalities. *Journal of Economic Dynamics & Control* **26**(6), pp. 889–909.

Dow, J. and Werlang, S. 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica* **60**, pp. 197–204.

El Ghaoui, L., Oustry, H. and Lebret, H. 1998. Robust solutions to uncertain semidefinite programs. *SIAM Journal of Optimization* **9**(1), pp. 33–52.

Ellsberg, D. 1961. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* **75**(4), pp. 643–669.

Epstein, L. and Wang, T. 1994. Intertemporal asset pricing under Knightian uncertainty. *Econometrica* **62**(2), pp. 283–322.

Fama, E. and French, K. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* **96**, pp. 246–273.

Gilboa, I. and Schmeidler, D. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* **18**(2), pp. 141–153.

Goldfarb, D. and Iyengar, G. 2003. Robust portfolio selection problems. *Mathematics of Operations Research* **28**(1), pp. 1–38.

Hansen, L., Sargent, T., Turmuhambetova, G. and Williams, N. 2002. Robust control and model uncertainty. *University of Chicago and Stanford University.* Working Paper.

Huber, P.: 1977. *Robust statistical procedures*. SIAM. Philadelphia.

Huber, P.: 1981. *Robust statistics*. Wiley. New York.

Jagannathan, R. and Ma, T. 2003. Risk reduction in large portfolios – why imposing wrong constraints helps. *Journal of Finance* **58**(4), pp. 1651–1684.

James, W. and Stein, C.: 1961. Estimation with quadratic loss. *Proceedings of the fourth Berkely symposium on Probability and Statistics*. Berkeley. University of California Press. pp. 361–379.

Jegadeesh, N. and Titman, S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* **48**, pp. 65–91.

Jobson, J. and Korkie, B. 1981. Putting Markowitz theory to work. *Journal of Portfolio Management* **7**(4), pp. 70–74.

Jobson, J., Korkie, B. and Ratti, V. 1979. Improved estimation for Markowitz portfolio using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section* pp. 279–284.

Jorion, P. 1986. Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* **21**(3), pp. 279–292.

Judge, G. and Bock, M.: 1978. *The statistical implications of pre-test and Stein-rule estimators in econometrics*. Wiley. New York.

Kandel, S. and Stambaugh, R. 1996. On the predictability of stoch returns: an asset-allocation perspective. *The Journal of Finance* **51**, pp. 385–424.

Karmarkar, N. 1984. A new polynomial–time algorithm for linear programming. *Proceedings of the 16th Annual ACM Symposium on Theory of Computing* pp. 302–311.

Khachiyan, L. 1979. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady* **20**, pp. 191–194.

Knight, F.: 1921. *Risk, uncertainty, and profit*. PhD thesis. Boston: Houghton Mifflin.

Kogan, L. and Wang, T.: 2002. A simple theory of asset pricing under model uncertainty. Working paper, University of British Columbia.

Ledoit, O.: 1994. Portfolio selection: improved covariance matrix estimation. Working Paper, Sloan school of management.

Lintner, J. 1965. The valuation of risk assets and selection of risky investment in stock portfolios and capital budgets. *Review of Economics and Statistics* **47**, pp. 13–37.

Lobo, M.: 2000. *Robust and convex optimization with applications in finance*. PhD thesis. Stanford University.

Luo, Z., Sturm, J. and Zhang, S.: 2001. Cones of nonnegative mappings – complexity and applications. Working Paper, CentER Tilburg University.

Lütkepohl, H.: 1996. *Handbook of matrices*. John Wiley & Sons.

Maenhout, P.: 1999. *Robust portfolio rules and asset pricing.* PhD thesis. Harvard University.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* **7**(1), pp. 77–91.

Merton, R. 1973. An intertemporal asset pricing model. *Econometrica* **41**, pp. 867–887.

Michaud, R. 1989. The Markowitz optimization enigma: is 'optimized' optimal?. *Financial Analyst Journal* **45**(1), pp. 31–42.

Michaud, R.: 1998. *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation.* Harvard Business School Press.

Mossin, J. 1966. Equilibrium in capital asset market. *Econometrica* **34**, pp. 768–783.

Nesterov, Y. and Nemirovski, A.: 1994. *Interior-point polynomial algorithms in convex programming.* SIAM. Philadelphia.

Nilim, A. and El Ghaoui, L.: 2002. Robust solutions to Markov decision problems. Presentation at NSF-ADP Workshop.

Pàstor, L. 2000. Portfolio selection and asset pricing models. *Journal of Finance* **55**, pp. 179–223.

Pàstor, L. and Stambaugh, F. 2000. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* **56**, pp. 335–381.

Pathak, P. 2002. Notes on robust portfolio choice. *Discussion paper, Harvard University.*

Rustem, B., Becker, R. and Marty, W. 2000. Robust min-max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control* **24**, pp. 1591–1621.

Savage, S.: 1954. *The foundations of statistics.* Wiley. New York.

Sharpe, W. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* **19**, pp. 425–442.

Sims, C. 2001. Pitfalls of minimax approaches to model uncertainty. *American Economic Review* **91**, pp. 51–54.

Sturm, J. 1999. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software* **11–12**, pp. 625–653.

Sturm, J. and Zhang, S. 2003. On cones of nonnegative quadratic functions. *Mathematics of Operations Research* **28**(2), pp. 246–267.

Uppal, R. and Wang, T. 2003. Model misspecification and underdiversification. *Journal of Finance* **58**(6), pp. 2465–2486.

Xia, Y. 2001. Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation.. *Journal of Finance* **56**(1), pp. 205–246.

Yakubovich, V. 1977. S-procedure in nonlinear control theory. *Vestnik Leningrad University* **4**, pp. 73–93.

Nederlandse Samenvatting

De toegevoegde waarde van dit proefschrift, *Robuuste portefeuilleoptimalisatie*, bestaat uit de ontwikkeling van modellen voor robuuste portefeuillekeuze, het uitwerken van rekenkundige methodieken ter optimalisatie van robuuste portefeuilleproblemen en een studie van het empirisch beleggingsresultaat van een robuuste aanpak van praktisch relevante problemen.

Robuust verwijst naar een benadering van onzekerheid omtrent de modelspecificatie en bijbehorende parameterwaarden die samen de rendementen beschrijven. Uitgangspunt is dat het rendementsmodel, i.e. de modelspecificatie en bijbehorende parameters, niet precies bekend is. In deze context bestuderen wij de besluitvorming van een rationele belegger die de portefeuillekeuze baseert op een analyse van rendementen maar onvoldoende informatie heeft om het rendementsmodel vast te stellen. De belegger weet slechts dat het rendementsmodel een element is van een verzameling van modellen. De 'robuuste belegger' streeft naar een oplossing met het beste beleggingsresultaat voor het slechtste geval dat mogelijk is binnen de gegeven verzameling van rendementsmodellen.

Robuuste optimalisatie is pas recentelijk bekeken in een financieringscontext. Dit is verrassend gezien de beloning die een verbeterde investeringsstrategie kan opleveren. Ondanks de enorme aandacht voor het ontwikkelen van modellen ter verklaring van rendementen zijn er maar weinig technieken die de beschikbare informatie en de bijbehorende onzekerheid naar een betrouwbare beslissing vertalen. Het probleem is dat optimale beslissingen veelal gevoelig zijn voor kleine veranderingen in de informatie, in dit geval het rendementsmodel. Aan de andere kant zijn onnauwkeurigheden in het model onvermijdelijk in een onzekere wereld. Zodoende zal een naïeve belegger - dat wil zeggen een belegger die onzekerheid negeert en een specifiek rendementsmodel hanteert - door eventuele onnauwkeurigheden in dat specifieke model suboptimale beslissingen nemen: de portefeuillekeuze is weliswaar optimaal onder het specifieke rendementsmodel maar is allerminst gunstig in werkelijkheid.

Cruciaal voor een robuuste aanpak van portefeuilleproblemen is de verzameling van rendementsmodellen die wordt meegenomen in de beslissing. De samenstelling van deze verzameling bepaalt namelijk de mate van robuustheid. Wij gebruiken statistische (Bayesiaanse) technieken die, gegeven de data, de belegger voorzien van een verzameling van aannemelijke rendementsmodellen. De robuuste belegger neemt alle aannemelijke modellen mee en is vol vertrouwen dat deze verzameling ook het toepasselijke rendementsmodel bevat. Derhalve is hij ervan overtuigd dat het robuust verwacht beleggingsresultaat ook daadwerkelijk zal worden behaald.

Men kan zich voorstellen dat een belegger terughoudend is over een robuuste aanpak: een verbetering in het bereiken van de ene doelstelling, in dit geval een robuust beleggingsresultaat, gaat doorgaans ten koste van het realiseren van een andere doelstelling, in dit geval het verwachte rendement. Echter, onder onzekerheid is de beslissing die voor een maximaal rendement zorgt onbekend en leidt een naïeve aanpak bijgevolg tot een slechte portefeuillekeuze. Een robuuste aanpak daarentegen houdt expliciet rekening met onzekerheid en kan in dit geval een waardevol alternatief zijn.

Hoofdstuk 1 vormt de inleiding tot robuuste portefeuilleoptimalisatie. We beschrijven het verband met andere benaderingen van beslissen onder onzekerheid. Tevens geven we een technische introductie tot robuuste optimalisatie. Een robuust optimalisatieprobleem is moeilijker op te lossen dan het originele, niet-robuuste optimalisatieprobleem. De mogelijkheid om de problemen op te lossen berust op een transformatie van het robuust (veelal oneindig-dimensionaal) probleem naar een equivalent eindig-dimensionaal probleem. We geven enkele voorbeelden van dit soort transformaties. Deze technieken worden in de verdere hoofdstukken gebruikt.

In hoofdstuk 2 bekijken we een robuuste aanpak van eenperiodeportefeuillekeuze onder schattingsonzekerheid. Voor portefeuillekeuze met mean-variance preferenties als ook met een restrictie op de maximale afwijking tot een benchmark leiden we het robuust optimalisatiemodel af. Bij de robuuste portefeuillekeuze verloopt de besluitvorming als volgt: 1) De robuuste aanpak test of de markt voldoende betrouwbare investeringsmogelijkheden biedt. 2) Als de markt de test doorstaat dan investeert de robuuste belegger in risicovolle titels, maar wel minder dan een naïeve belegger. 3) De robuuste belegger gebruikt een portefeuillecompositie die afwijkt van de portefeuillecompositie die volgt uit een naïve aanpak als de maatstaf voor onzekerheid afwijkt van de risicomaatstaf. Voor enkele, voor de praktijk relevante, portefeuilleproblemen kunnen we de robuuste portefeuille analytisch uitdrukken. Voor het algemene geval kunnen we de robuuste portefeuille efficiënt numeriek berekenen. Dit stelt ons in staat een empirische studie van de robuuste aanpak te doen. De resultaten van deze studie laten zien dat 'schattingsrobuuste' portefeuilles een betrouwbaar beleggingsresultaat leveren dat significant beter is dan het beleggingsresultaat onder een naïeve aanpak.

In hoofdstuk 3 bekijken we naast schattingsonzekerheid ook onzekerheid in de modelspecificatie. Het uitgangspunt is dat financiële experts de belegger verschillende, mogelijk conflicterende modelspecificaties zullen adviseren. De modelrobuuste belegger neemt elk van deze modelspecificaties in overweging en tracht het beleggingsresultaat voor de minst wenselijke modelspecificatie te maximaliseren. Analytische resultaten laten zien dat de modelrobuuste portefeuille is gebaseerd op een - endogeen bepaalde - convexe combinatie van deze verschillende modelspecificaties. Deze resultaten laten ook zien dat een significante verbetering van betrouwbaarheid kan worden behaald tegen een geringe daling, zoniet een stijging, van het verwachte beleggingsresultaat. Voor relevante situaties is het daadwerkelijke (ex-post) beleggingsresultaat van een robuuste portefeuille zelfs beter dan het beleggingsresultaat van elk van de portefeuilles gebaseerd op een van de verschillende modelspecificaties. Deze analytische resultaten worden bevestigd door een empirische test van de modelrobuuste aanpak waarbij we enkele vooraanstaande gestructureerde modellen, zoals het CAPM en het Fama & French model, meenemen. De resultaten laten ook zien dat naast modelrobuustheid, schattingsrobuustheid zeker zo belangrijk is voor een goed beleggingsresultaat onder onzekerheid. Een schattingsrobuuste aanpak gebaseerd op een ongestructureerd model evenaart het beleggingsresultaat van de gestructureerde modellen.

Hoofdstuk 4 betreft meerperiodeportefeuillekeuze onder schattingsonzekerheid. We bekijken een buy-and-hold belegger en een belegger die elke jaar de portefeuille aanpast. Beide beleggers hanteren een model met voorspelbaarheid in de rendementen. Het doel is de robuuste portefeuille te vergelijken met naïeve portefeuilles en met de portefeuillekeuze die volgt uit een Bayesiaanse aanpak van onzekerheid. De combinatie van onzekerheid, het dynamisch rendementsmodel en de doelfunctie beletten een transformatie naar een eenvoudig optimalisatieprobleem. Daarom berekenen we de robuuste portefeuillekeuze door middel van numerieke methodes. De resultaten tonen dat de robuuste belegger aanzienlijk minder investeert in risicovolle activa dan een naïeve of Bayesiaanse belegger. Als de onzekerheid groot is dan investeert de robuuste belegger zelfs helemaal niet in risicovolle activa. Een interessant bijproduct van de robuuste portefeuilles is de karakterisatie van de minst wenselijke, aannemelijke parameterconfiguratie voor portefeuillekeuze. Deze geeft aan welke eigenschappen van het model cruciaal zijn voor de portefeuillekeuze. Afhankelijk van de initiële economische situatie (dividend) en de investeringshorizon zijn persistentie van een hoog initieel dividend, de langetermijnvoorspelling van het dividend en de voorspelkracht van dividend van belang.

In hoofdstuk 5 bekijken we robuuste eenperiodeportefeuillekeuze met opties. Door de introductie van opties zijn de transformatietechnieken voor robuuste problemen zoals in hoofdstukken 2 en 3 niet toepasbaar. We ontwikkelen een equivalente formulering voor het robuuste probleem met eindige dimensies dat efficiënt oplosbaar is. We illustreren de techniek aan de hand van een benchmark tracking probleem.

De conclusie (hoofdstuk 6) is dat - in een onzekere wereld - een robuuste benadering van portefeuillekeuze tot een aanzienlijk beter beleggingsresultaat leidt dan een benadering die onzekerheid negeert.

Curriculum Vitae

Frank Lutgens was born on August 23, 1977 in Sittard, The Netherlands. He studied econometrics, main subject operations research, at Maastricht University. His masters thesis, "IP network design", was researched and written during an internship at KPN Research in Leidschendam, the Netherlands. For this thesis, he was rewarded the student-research award of the UM. He obtained his MSc in econometrics with distinction in 1999.

From 1999 to 2003 Frank Lutgens worked on his Ph.D. dissertation 'Robust Portfolio Optimization' at the Department of Quantitative Economics of Maastricht University. During this period he obtained the certificate of the Dutch Network of Mathematical Operations Research (LNMB). He has been involved in organizing, lecturing and tutoring courses for the Departments of Finance and Quantitative Economics. In 2003 a research proposal proceeding from the dissertation was awarded a grant from Inquire Europe. Since January 1, 2004 he has been working as a researcher for the Finance Department, Maastricht University. In March 2004 he participated as an economic expert in the economic technical committee associated with the United Nations negotiations in Cyprus.